B13

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 March 2005 (31.03.2005)

PCT

(10) International Publication Number
WO 2005/027980 A1

(51) International Patent Classification[7]: A61K 48/00, C07H 21/04

(21) International Application Number: PCT/US2004/029968

(22) International Filing Date: 13 September 2004 (13.09.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/502,678 12 September 2003 (12.09.2003) US

(71) Applicant *(for all designated States except US)*: **UNIVERSITY OF MASSACHUSETTS** [US/US]; 365 Plantation Street, Worcester, MA 01605 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ARONIN, Neil** [US/US]; 19 Whittier Road, Newtonville, MA 02460 (US). **ZAMORE, Phillip, D.** [US/US]; 500 Green Street, Northboro, MA 01532 (US).

(74) Agents: **MILASINCIC, Debra, J.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: RNA INTERFERENCE FOR THE TREATMENT OF GAIN-OF-FUNCTION DISORDERS

(57) Abstract: The present invention relates to the discovery of an effective treatment for a variety of gain-of-function diseases, in particular, Huntington's disease (HD). The present invention utilizes RNA Interference technology (RNAi) against polymorphic regions in the genes encoding various gain-of-function mutant proteins resulting in an effective treatment for the gain-of-function disease.

WO 2005/027980 A1

RNA INTERFERENCE FOR THE TREATMENT OF GAIN-OF-FUNCTION DISORDERS

Related Applications

This patent application claims the benefit of U.S. Provisional Patent Application Serial No. 60/502,678, entitled "RNA Interference for the Treatment of Gain-of-Function Disorders", filed September 12, 2003. The entire contents of the above-referenced provisional patent applications are incorporated herein by this reference.

Background of the Invention

RNA interference (RNAi) is the mechanism of sequence-specific, post-transcriptional gene silencing initiated by double-stranded RNAs (dsRNA) homologous to the gene being suppressed. dsRNAs are processed by Dicer, a cellular ribonuclease III, to generate duplexes of about 21 nt with 3'-overhangs (small interfering RNA, siRNA) which mediate sequence-specific mRNA degradation. In mammalian cells siRNA molecules are capable of specifically silencing gene expression without induction of the unspecific interferon response pathway. Thus, siRNAs have become a new and powerful alternative to other genetic tools such as antisense oligonucleotides and ribozymes to analyze gene function. Moreover, siRNA's are being developed for therapeutic purposes with the aim of silencing disease genes in humans.

Trinucleotide repeat diseases comprise a recently recognized group of inherited disorders. The common genetic mutation is an increase in a series of a particular trinucleotide repeat. To date, the most frequent trinucleotide repeat is CAG, which codes for the amino acid glutamine. At least 9 CAG repeat diseases are known and there are more than 20 varieties of these diseases, including Huntington's disease, Kennedy's disease and many spinocerebellar diseases. These disorders share a neurodegenerative component in the brain and/or spinal cord. Each disease has a specific pattern of neurodegeneration in the brain and most have an autosomal dominant inheritance.

The onset of the diseases generally occurs at 30 to 40 years of age, but in Huntington's disease CAG repeats in the huntingtin gene of >60 portend a juvenile onset.

Recent research by the instant inventors has shown that the genetic mutation (increase in length of CAG repeats from normal <36 in the huntingtin gene to >36 in

disease) is associated with the synthesis of a mutant huntingtin protein, which has >36 polyglutamines (Aronin et al., 1995). It has also been shown that the protein forms cytoplasmic aggregates and nuclear inclusions (Difiglia et al., 1997) and associates with vesicles (Aronin et al., 1999). The precise pathogenic pathways are not known.

Huntington's disease (and by implication other trinucleotide repeat diseases) is believed to be caused, at least in part, by aberrant protein interactions, which cause impairment of critical neuronal processes, neuronal dysfunction and ultimately neuronal death (neurodegeneration in brain areas called the striatum and cortex). In the search for an effective treatment for these diseases, researchers in this field emphasized understanding the pathogenesis of the disease and initially sought to intercede at the level of the presumed aberrant protein interactions. However, there is no effective treatment for Huntington's disease or other trinucleotide repeat diseases. Moreover, it is now appreciated that multiple abnormal processes might be active in these types of disease.

Summary of the Invention

The present invention relates to the methods for treating a variety of gain-of-function diseases. In particular, the invention provides methods for the selective destruction of mutant mRNAs transcribed from gain-of-function mutant genes, thus preventing production of the mutant proteins encoded by such genes. Other RNAi-based methods for destroying mutant genes have been proposed in which siRNAs are targeted to, for example, a point mutation occurring in a single allele in the mutant gene (e.g., the point mutation in the superoxide dismutase (SOD) gene associated with amyotrophic lateral sclerosis (ALS)). However, there is a key difference between ALS and trinucleotide repeat diseases, such as Huntington's disease. ALS has a point mutation in one allele as the genetic change whereas trinucleotide repeat diseases have an expanded CAG repeat region in one allele as the genetic change. Use of RNAi against the expanded CAG repeat region has potential complications. Over 80 normal genes with CAG repeat regions are known to exist in cells. Thus, siRNAs targeting these CAG repeats cannot be used without risking widespread destruction of normal CAG repeat-containing mRNAs. Likewise, targeting non-allele-specific sites would result in loss of both normal and mutant huntingtin causes neuronal dysfunction.

The methods of the invention utilize RNA interference technology (RNAi) against selected polymorphic regions (*i.e.*, regions containing allele-specific or allelic polymorphisms) which are distinct from the site of mutation in the genes encoding mutant proteins. The methodologies of the instant invention are effective treatments for gain-of-function diseases resulting from deletion mutations, insertion mutations, point mutations, and the like, provided that the mutant gene encodes a protein having a function not normally associated with wild type protein.

In a preferred aspect, the methodologies of the instant invention provide an effective treatment for Huntington's disease (HD). The methodologies also provide effective treatments for other polyglutamine disorders and/or trinucleotide repeat disease, as described in detail herein.

Accordingly, in one aspect, the present invention provides a method of treating a subject having or at risk of having a disease characterized or caused by a gain of function mutant protein by administering to the subject an effective amount of an RNAi agent targeting an allelic polymorphism within a gene encoding a mutant protein e.g.,) huntingtin protein, such that sequence-specific interference of a gene occurs resulting in an effective treatment for the disease. In one embodiment, the mutant protein contains an expanded polyglutamine region. In another one embodiment, the gene encoding the mutant protein contains an expanded trinucleotide repeat region.

In a yet another embodiment, the method of the invention can be used to treat Huntington's disease and a variety of other diseases selected from the group consisting of spino-cerebellar ataxia type 1, spino-cerebellar ataxia type 2, spino-cerebellar ataxia type 3, spino-cerebellar ataxia type 6, spino-cerebellar ataxia type 7, spino-cerebellar ataxia type 8, spino-cerebellar ataxia type 12, myotonic dystrophy, spinal bulbar muscular disease and dentatoiubral-pallidoluysian atrophy.

The method of the invention uses RNAi agents homologous to an allelic polymorphism within the gene encoding, for example, a mutant huntingtin protein for the treatment of Huntington's disease. In a preferred embodiment, the RNAi agent targets allelic polymorphism selected from the group consisting of P1-P5. In a further preferred embodiment, the RNAi agent targets an allelic polymorphism selected from the group consisting of P6-P43.

In a further embodiment, the invention provides RNAi agents comprising of a first and second strand each containing 16-25 nucleotides. The first strand of the present invention is homologous to a region of a gene encoding a gain-of-function mutant protein, wherein the nucleotide sequence of the gain-of-function mutant protein comprises an allelic polymorphism. The second strand includes 16-25 nucleotides complementary to the first strand. The RNAi agent can also have a loop portion comprising 4-11, e.g., 4, 5, 6, 7, 8, 9, 10, 11, nucleotides that connects the two nucleotides sequences. In still other embodiments, the target region of the mRNA sequence is located in a 5' untranslated region (UTR) or a 3' UTR of the mRNA of a mutant protein.

In another embodiment, the invention provides an expression construct comprising an isolated nucleic acid that encodes a nucleic acid molecule with a first sequence of 16-25 nucleotides homologous to an allelic polymorphism within, for example, the gene encoding a mutant huntingtin protein. The expression construct can be for example, a viral vector, retroviral vector, expression cassette or plasmid. The expression construct can also have an RNA polymerase II promoter sequence or RNA Polymerase II promoter sequence, such as, U6 snRNA promoter of H1 promoter.

In yet other embodiments, the present invention provides host cells e.g.,) mammalian cells) comprising nucleic acid molecules and expression constructs of the present invention.

In still other embodiments, the present invention provides therapeutic compositions comprising the nucleic acid molecules of the invention and a pharmaceutically acceptable carrier.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1a-k: Human huntingtin gene, nucleotide sequence (SEQ ID NO:1)

Figure 2a-b: Human huntingtin protein, amino acid sequence (SEQ ID NO:2)

Figure 3: Sense (SEQ ID NO: 3) and antisense (SEQ ID NO: 4) of the huntingtin (htt) target RNA sequence

Figure 4: Thermodynamic analysis of siRNA strand 5' ends for the siRNA duplex

Figure 5a-c: *In vitro* RNAi reactions programmed with siRNA targeting a polymorphism within the huntingtin (htt) mRNA. (a) Standard siRNA. (b) siRNA improved by reducing the base-pairing strength of the 5′ end of the anti-sense strand of the siRNA duplex. (c) siRNA improved by reducing the unpairing the 5′ end of the anti-sense strand of the siRNA duplex.

Figure 6a-b. RNAi of endogenous Htt protein in HeLa cells. (a) Immunoblot of human Htt protein. (b) Quantification of same.

Detailed Description of the Invention

The present invention relates to methods and reagents for treating a variety of gain-of-function diseases. In one aspect, the invention relates to methods and reagents for treating a variety of diseases characterized by a mutation in one allele or copy of a gene, the mutation encoding a protein which is sufficient to contribute to or cause the disease. Preferably, the methods and reagents are used to treat diseases caused or characterized by a mutation that is inherited in an autosomal dominant fashion. In one embodiment, the methods and reagents are used for treating a variety of neurodegenerative disease caused by a gain-of-function mutation, e.g., polyglutamine disorders and/or trinucleotide repeat diseases, for example, Huntington's disease. In another embodiment, the methods and reagents are used for treating diseases caused by a gain-of-function in an oncogene, the mutated gene product being a gain-of-function mutant, e.g., cancers caused by a mutation in the *ret* oncogene (e.g., *ret-1*), for example, endocrine tumors, medullary thyroid tumors, parathyroid hormone tumors, multiple endocrine neoplasia type2, and the like. In another embodiment, the methods and reagents of the invention can be used to treat a variety of gastrointestinal cancers known to be caused by an autosomally-inherited, gain-of-function mutations.

The present invention utilizes RNA interference technology (RNAi) against allelic polymorphisms located within a gene encoding a gain-of-function mutant protein. RNAi destroys the corresponding mutant mRNA with nucleotide specificity and selectivity. RNA agents of the present invention are targeted to polymorphic regions of a mutant gene, resulting in cleavage of mutant mRNA. These RNA agents, through a series of protein-nucleotide interactions, function to cleave the mutant mRNAs. Cells

destroy the cleaved mRNA, thus preventing synthesis of corresponding mutant protein e.g., the huntingtin protein.

Accordingly, in one aspect, the present invention provides a method of treating a subject having or at risk of having a disease characterized or caused by a gain of function mutant protein by administering to the subject an effective amount of an RNAi agent targeting an allelic polymorphism within a gene encoding a mutant protein e.g.,) huntingtin protein, such that sequence-specific interference of a gene occurs resulting in an effective treatment for the disease. In one embodiment, the mutant protein contains an expanded polyglutamine region. In another one embodiment, the gene encoding the mutant protein contains an expanded trinucleotide repeat region.

In a yet another embodiment, the method of the invention can be used to treat Huntington's disease and a variety of other diseases selected from the group consisting of spino-cerebellar ataxia type 1, spino-cerebellar ataxia type 2, spino-cerebellar ataxia type 3, spino-cerebellar ataxia type 6, spino-cerebellar ataxia type 7, spino-cerebellar ataxia type 8, spino-cerebellar ataxia type 12, myotonic dystrophy, spinal bulbar muscular disease and dentatoiubral-pallidoluysian atrophy.

The method of the invention uses RNAi agents homologous to an allelic polymorphism within the gene encoding, for example, a mutant huntingtin protein for the treatment of Huntington's disease. In a preferred embodiment, the RNAi agent targets allelic polymorphism selected from the group consisting of P1-P5. In a further preferred embodiment, the RNAi agent targets an allelic polymorphism selected from the group consisting of P6-P43.

In a further embodiment, the invention provides RNAi agents comprising of a first and second strand each containing 16-25 nucleotides. The first strand of the present invention is homologous to a region of a gene encoding a gain-of-function mutant protein, wherein the nucleotide sequence of the gain-of-function mutant protein comprises an allelic polymorphism. The second strand includes 16-25 nucleotides complementary to the first strand. The RNAi agent can also have a loop portion comprising 4-11, e.g., 4, 5, 6, 7, 8, 9, 10, 11, nucleotides that connect the two nucleotides sequences. In still other embodiments, the target region of the mRNA sequence is located in a 5' untranslated region (UTR) or a 3' UTR of the mRNA of a mutant protein.

In another embodiment, the invention provides an expression construct comprising an isolated nucleic acid that encodes a nucleic acid molecule with a first sequence of 16-25 nucleotides homologous to an allelic polymorphism within, for example, the gene encoding a mutant huntingtin protein. The expression construct can be for example, a viral vector, retroviral vector, expression cassette or plasmid. The expression construct can also have an RNA polymerase II promoter sequence or RNA Polymerase II promoter sequence, such as, U6 snRNA promoter of H1 promoter.

In yet other embodiments, the present invention provides host cells e.g.,) mammalian cells) comprising nucleic acid molecules and expression constructs of the present invention.

In still other embodiments, the present invention provides therapeutic compositions comprising the nucleic acid molecules of the invention and a pharmaceutically acceptable carrier.

So that the invention may be more readily understood, certain terms are first defined.

The term "nucleoside" refers to a molecule having a purine or pyrimidine base covalently linked to a ribose or deoxyribose sugar. Exemplary nucleosides include adenosine, guanosine, cytidine, uridine and thymidine. Additional exemplary nucleosides include inosine, 1-methyl inosine, pseudouridine, 5,6-dihydrouridine, ribothymidine, ^{2}N-methylguanosine and ^{2,2}N,N-dimethylguanosine (also referred to as "rare" nucleosides). The term "nucleotide" refers to a nucleoside having one or more phosphate groups joined in ester linkages to the sugar moiety. Exemplary nucleotides include nucleoside monophosphates, diphosphates and triphosphates. The terms "polynucleotide" and "nucleic acid molecule" are used interchangeably herein and refer to a polymer of nucleotides joined together by a phosphodiester linkage between 5' and 3' carbon atoms.

The term "RNA" or "RNA molecule" or "ribonucleic acid molecule" refers to a polymer of ribonucleotides. The term "DNA" or "DNA molecule" or deoxyribonucleic acid molecule" refers to a polymer of deoxyribonucleotides. DNA and RNA can be synthesized naturally (e.g., by DNA replication or transcription of DNA, respectively).

RNA can be post-transcriptionally modified. DNA and RNA can also be chemically synthesized. DNA and RNA can be single-stranded (*i.e.*, ssRNA and ssDNA, respectively) or multi-stranded (e.g., double stranded, *i.e.*, dsRNA and dsDNA, respectively). "mRNA" or "messenger RNA" is single-stranded RNA that specifies the amino acid sequence of one or more polypeptide chains. This information is translated during protein synthesis when ribosomes bind to the mRNA.

As used herein, the term "small interfering RNA" ("siRNA") (also referred to in the art as "short interfering RNAs") refers to an RNA (or RNA analog) comprising between about 10-50 nucleotides (or nucleotide analogs) which is capable of directing or mediating RNA interference. Preferably, a siRNA comprises between about 15-30 nucleotides or nucleotide analogs, more preferably between about 16-25 nucleotides (or nucleotide analogs), even more preferably between about 18-23 nucleotides (or nucleotide analogs), and even more preferably between about 19-22 nucleotides (or nucleotide analogs) (e.g., 19, 20, 21 or 22 nucleotides or nucleotide analogs). The term "short" siRNA refers to a siRNA comprising ~21 nucleotides (or nucleotide analogs), for example, 19, 20, 21 or 22 nucleotides. The term "long" siRNA refers to a siRNA comprising ~24-25 nucleotides, for example, 23, 24, 25 or 26 nucleotides. Short siRNAs may, in some instances, include fewer than 19 nucleotides, e.g., 16, 17 or 18 nucleotides, provided that the shorter siRNA retains the ability to mediate RNAi. Likewise, long siRNAs may, in some instances, include more than 26 nucleotides, provided that the longer siRNA retains the ability to mediate RNAi absent further processing, e.g., enzymatic processing, to a short siRNA.

The term "nucleotide analog" or "altered nucleotide" or "modified nucleotide" refers to a non-standard nucleotide, including non-naturally occurring ribonucleotides or deoxyribonucleotides. Preferred nucleotide analogs are modified at any position so as to alter certain chemical properties of the nucleotide yet retain the ability of the nucleotide analog to perform its intended function. Examples of positions of the nucleotide which may be derivitized include the 5 position, e.g., 5-(2-amino)propyl uridine, 5-bromo uridine, 5-propyne uridine, 5-propenyl uridine, etc.; the 6 position, e.g., 6-(2-amino)propyl uridine; the 8-position for adenosine and/or guanosines, e.g., 8-bromo guanosine, 8-chloro guanosine, 8-fluoroguanosine, etc. Nucleotide analogs also include deaza nucleotides, e.g., 7-deaza-adenosine; O- and N-modified (e.g., alkylated, e.g., N6-methyl adenosine, or as otherwise known in the art) nucleotides; and other

heterocyclically modified nucleotide analogs such as those described in Herdewijn, Antisense Nucleic Acid Drug Dev., 2000 Aug. 10(4):297-310.

Nucleotide analogs may also comprise modifications to the sugar portion of the nucleotides. For example the 2' OH-group may be replaced by a group selected from H, OR, R, F, Cl, Br, I, SH, SR, NH_2, NHR, NR_2, COOR, or OR, wherein R is substituted or unsubstituted C_1 –C_6 alkyl, alkenyl, alkynyl, aryl, etc. Other possible modifications include those described in U.S. Patent Nos. 5,858,988, and 6,291,438.

The phosphate group of the nucleotide may also be modified, e.g., by substituting one or more of the oxygens of the phosphate group with sulfur (e.g., phosphorothioates), or by making other substitutions which allow the nucleotide to perform its intended function such as described in, for example, Eckstein, Antisense Nucleic Acid Drug Dev. 2000 Apr. 10(2):117-21, Rusckowski et al. Antisense Nucleic Acid Drug Dev. 2000 Oct. 10(5):333-45, Stein, Antisense Nucleic Acid Drug Dev. 2001 Oct. 11(5): 317-25, Vorobjev et al. Antisense Nucleic Acid Drug Dev. 2001 Apr. 11(2):77-85, and U.S. Patent No. 5,684,143. Certain of the above-referenced modifications (e.g., phosphate group modifications) preferably decrease the rate of hydrolysis of, for example, polynucleotides comprising said analogs in vivo or in vitro.

The term "oligonucleotide" refers to a short polymer of nucleotides and/or nucleotide analogs. The term "RNA analog" refers to an polynucleotide (e.g., a chemically synthesized polynucleotide) having at least one altered or modified nucleotide as compared to a corresponding unaltered or unmodified RNA but retaining the same or similar nature or function as the corresponding unaltered or unmodified RNA. As discussed above, the oligonucleotides may be linked with linkages which result in a lower rate of hydrolysis of the RNA analog as compared to an RNA molecule with phosphodiester linkages. For example, the nucleotides of the analog may comprise methylenediol, ethylene diol, oxymethylthio, oxyethylthio, oxycarbonyloxy, phosphorodiamidate, phophoroamidate, and/or phosphorothioate linkages. Preferred RNA analogues include sugar- and/or backbone-modified ribonucleotides and/or deoxyribonucleotides. Such alterations or modifications can further include addition of non-nucleotide material, such as to the end(s) of the RNA or internally (at one or more nucleotides of the RNA). An RNA analog need only be sufficiently similar to natural RNA that it has the ability to mediate (mediates) RNA interference.

As used herein, the term "RNA interference" ("RNAi") refers to a selective intracellular degradation of RNA. RNAi occurs in cells naturally to remove foreign RNAs (e.g., viral RNAs). Natural RNAi proceeds via fragments cleaved from free dsRNA which direct the degradative mechanism to other similar RNA sequences. Alternatively, RNAi can be initiated by the hand of man, for example, to silence the expression of target genes.

An RNAi agent having a strand which is "sequence sufficiently complementary to a target mRNA sequence to direct target-specific RNA interference (RNAi)" means that the strand has a sequence sufficient to trigger the destruction of the target mRNA by the RNAi machinery or process.

As used herein, the term "isolated RNA" (e.g., "isolated siRNA" or "isolated siRNA precursor") refers to RNA molecules which are substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

The term "in vitro" has its art recognized meaning, e.g., involving purified reagents or extracts, e.g., cell extracts. The term "in vivo" also has its art recognized meaning, e.g., involving living cells, e.g., immortalized cells, primary cells, cell lines, and/or cells in an organism.

As used herein, the term "transgene" refers to any nucleic acid molecule, which is inserted by artifice into a cell, and becomes part of the genome of the organism that develops from the cell. Such a transgene may include a gene that is partly or entirely heterologous (*i.e.*, foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism. The term "transgene" also means a nucleic acid molecule that includes one or more selected nucleic acid sequences, e.g., DNAs, that encode one or more engineered RNA precursors, to be expressed in a transgenic organism, e.g., animal, which is partly or entirely heterologous, *i.e.*, foreign, to the transgenic animal, or homologous to an endogenous gene of the transgenic animal, but which is designed to be inserted into the animal's genome at a location which differs from that of the natural gene. A transgene includes one or more promoters and any other DNA, such as introns, necessary for expression of the selected nucleic acid sequence, all operably linked to the selected sequence, and may include an enhancer sequence.

A gene "involved" in a disease or disorder includes a gene, the normal or aberrant expression or function of which effects or causes the disease or disorder or at least one symptom of said disease or disorder

The term "gain-of-function mutation" as used herein, refers to any mutation in a gene in which the protein encoded by said gene (*i.e.*, the mutant protein) acquires a function not normally associated with the protein (*i.e.*, the wild type protein) causes or contributes to a disease or disorder. The gain-of-function mutation can be a deletion, addition, or substitution of a nucleotide or nucleotides in the gene which gives rise to the change in the function of the encoded protein. In one embodiment, the gain-of-function mutation changes the function of the mutant protein or causes interactions with other proteins. In another embodiment, the gain-of-function mutation causes a decrease in or removal of normal wild-type protein, for example, by interaction of the altered, mutant protein with said normal, wild-type protein.

The term "polymorphism" as used herein, refers to a variation (e.g., a deletion, insertion, or substitution) in a gene sequence that is identified or detected when the same gene sequence from different sources subjects (but from the same organism) are compared. For example, a polymorphism can be identified when the same gene sequence from different subjects (but from the same organism) are compared. Identification of such polymorphisms is routine in the art, the methodologies being similar to those used to detect, for example, breast cancer point mutations. Identification can be made, for example, from DNA extracted from a subject's lymphocytes, followed by amplification of polymorphic regions using specific primers to said polymorphic region. Alternatively, the polymorphism can be identified when two alleles of the same gene are compared. A variation in sequence between two alleles of the same gene within an organism is referred to herein as an "allelic polymorphism". The polymorphism can be at a nucleotide within a coding region but, due to the degeneracy of the genetic code, no change in amino acid sequence is encoded. Alternatively, polymorphic sequences can encode a different amino acid at a particular position, but the change in the amino acid does not affect protein function. Polymorphic regions can also be found in non-encoding regions of the gene.

The term "polyglutamine domain," as used herein, refers to a segment or domain of a protein that consist of a consecutive glutamine residues linked to peptide bonds. In one embodiment the consecutive region includes at least 5 glutamine residues.

The term "expanded polyglutamine domain" or "expanded polyglutamine segment", as used herein, refers to a segment or domain of a protein that includes at least 35 consecutive glutamine residues linked by peptide bonds. Such expanded segments are found in subjects afflicted with a polyglutamine disorder, as described herein, whether or not the subject has shown to manifest symptoms.

The term "trinucleotide repeat" or "trinucleotide repeat region" as used herein, refers to a segment of a nucleic acid sequence e.g.,) that consists of consecutive repeats of a particular trinucleotide sequence. In one embodiment, the trinucleotide repeat includes at least 5 consecutive trinucleotide sequences. Exemplary trinucleotide sequences include, but are not limited to, CAG, CGG, GCC, GAA, CTG, and/or CGG.

The term "trinucleotide repeat diseases" as used herein, refers to any disease or disorder characterized by an expanded trinucleotide repeat region located within a gene, the expanded trinucleotide repeat region being causative of the disease or disorder. Examples of trinucleotide repeat diseases include, but are not limited to spino-cerebellar ataxia type 12 spino-cerebellar ataxia type 8, fragile X syndrome, fragile XE Mental Retardation, Friedreich's ataxia and myotonic dystrophy. Preferred trinucleotide repeat diseases for treatment according to the present invention are those characterized or caused by an expanded trinucleotide repeat region at the 5' end of the coding region of a gene, the gene encoding a mutant protein which causes or is causative of the disease or disorder. Certain trinucleotide diseases, for example, fragile X syndrome, where the mutation is not associated with a coding region may not be suitable for treatment according to the methodologies of the present invention, as there is no suitable mRNA to be targeted by RNAi. By contrast, disease such as Friedreich's ataxia may be suitable for treatment according to the methodologies of the invention because, although the causative mutation is not within a coding region (*i.e.*, lies within an intron), the mutation may be within, for example, an mRNA precursor (e.g., a pre-spliced mRNA precursor).

The term "polyglutamine disorder" as used herein, refers to any disease or disorder characterized by an expanded of a $(CAG)_n$ repeats at the 5' end of the coding region (thus encoding an expanded polyglutamine region in the encoded protein). In one embodiment, polyglutamine disorders are characterized by a progressive degeneration of nerve cells. Examples of polyglutamine disorders include but are not limited to: Huntington's disease, spino-cerebellar ataxia type 1, spino-cerebellar ataxia type 2, spino-cerebellar ataxia type 3 (also know as Machado-Joseph disease), and spino-

cerebellar ataxia type 6, spino-cerebellar ataxia type 7 and dentatoiubral-pallidoluysian atrophy.

The phrase "examining the function of a gene in a cell or organism" refers to examining or studying the expression, activity, function or phenotype arising therefrom.

Various methodologies of the instant invention include step that involves comparing a value, level, feature, characteristic, property, etc. to a "suitable control", referred to interchangeably herein as an "appropriate control". A "suitable control" or "appropriate control" is any control or standard familiar to one of ordinary skill in the art useful for comparison purposes. In one embodiment, a "suitable control" or "appropriate control" is a value, level, feature, characteristic, property, etc. determined prior to performing an RNAi methodology, as described herein. For example, a transcription rate, mRNA level, translation rate, protein level, biological activity, cellular characteristic or property, genotype, phenotype, etc. can be determined prior to introducing an RNAi agent of the invention into a cell or organism. In another embodiment, a "suitable control" or "appropriate control" is a value, level, feature, characteristic, property, etc. determined in a cell or organism, e.g., a control or normal cell or organism, exhibiting, for example, normal traits. In yet another embodiment, a "suitable control" or "appropriate control" is a predefined value, level, feature, characteristic, property, etc.

Various aspects of the invention are described in further detail in the following subsections.

I. Polyglutamine disorders

Polyglutamine disorders are a class of disease or disorders characterized by a common genetic mutation. In particular, the disease or disorders are characterized by an expanded repeat of the trinucleotide CAG which gives rise, in the encoded protein, to an expanded stretch of glutamine residues. Polyglutamine disorders are similar in that the diseases are characterized by a progressive degeneration of nerve cells. Despite their similarities, polyglutamine disorders occur on different chromosomes and thus occur on entirely different segments of DNA. Examples of polyglutamine disorders include

Huntington's disease, Dentatorubropallidoluysian Atrophy, Spinobulbar Muscular atrophy, Spinocerebellar Ataxia Type 1, Spinocerebellar Ataxia Type 2, Spinocerebellar Ataxia Type 3, Spinocerebellar Ataxia Type 6 and Spinocerebellar Ataxia Type 7 (Table 3).

5

Table 1. Polyglutamine disorders

Disease	Gene	Locus	Protein	CAG repeat size	
				Normal	Disease
Spinobulbar muscular atrophy (Kennedy disease)	*AR*	Xq13–21	Androgen receptor (AR)	9–36	38–62
Huntington's disease	*HD*	4p16.3	Huntingtin	6–35	36–121
Dentatorubral-pallidoluysian atrophy (Haw–River syndrome)	*DRPLA*	12p13.31	Atrophin-1	6–35	49–88
Spinocerebellar ataxia type 1	*SCA1*	6p23	Ataxin-1	6–44[a]	39–82
Spinocerebellar ataxia type 2	*SCA2*	12q24.1	Ataxin-2	15–31	36–63
Spinocerebellar ataxia type 3 (Machado–Joseph disease)	*SCA3* (*MJD1*)	14q32.1	Ataxin-3	12–40	55–84
Spinocerebellar ataxia type 6	*SCA6*	19p13	α_{1A}-voltage-dependent calcium channel subunit	4–18	21–33
Spinocerebellar ataxia type 7	*SCA7*	13p12–13	Ataxin-7	4–35	37–306

[a]Alleles with 21 or more repeats are interrupted by 1–3 CAT units; disease alleles contain pure CAG tracts.

Polyglutamine disorders of the invention are characterized by (e.g., domains having between about 30 to 35 glutamine residues, between about 35 to 40 glutamine residues, between about 40 to 45 glutamine residues and having about 45 or more glutamine residues. The polyglutamine domain typically contains consecutive glutamine residues (Q n>36).

II. Huntington Disease

Huntington's disease, inherited as an autosomal dominant disease, causes impaired cognition and motor disease. Patients can live more than a decade with severe debilitation, before premature death from starvation or infection. The disease begins in the fourth or fifth decade for most cases, but a subset of patients manifest disease in teenage years. The genetic mutation for Huntington's disease is a lengthened CAG repeat in the huntingtin gene. CAG repeat varies in number from 8 to 35 in normal individuals (Kremer et al., 1994). The genetic mutation e.g.,) an increase in length of the CAG repeats from normal less than 36 in the huntingtin gene to greater than 36 in the disease is associated with the synthesis of a mutant huntingtin protein, which has greater than 36 polyglutamates (Aronin et al., 1995). In general, individuals with 36 or more CAG repeats will get Huntington's disease. Prototypic for as many as twenty other diseases with a lengthened CAG as the underlying mutation, Huntington's disease still has no effective therapy. A variety of interventions -- such as interruption of apoptotic pathways, addition of reagents to boost mitochondrial efficiency, and blockade of NMDA receptors -- have shown promise in cell cultures and mouse model of Huntington's disease. However, at best these approaches reveal a short prolongation of cell or animal survival.

Huntington's disease complies with the central dogma of genetics: a mutant gene serves as a template for production of a mutant mRNA; the mutant mRNA then directs synthesis of a mutant protein (Aronin et al., 1995; DiFiglia et al., 1997). Mutant huntingtin (protein) probably accumulates in selective neurons in the striatum and cortex, disrupts as yet determined cellular activities, and causes neuronal dysfunction and death (Aronin et al., 1999; Laforet et al., 2001). Because a single copy of a mutant gene suffices to cause Huntington's disease, the most parsimonious treatment would render the mutant gene ineffective. Theoretical approaches might include stopping gene

transcription of mutant huntingtin, destroying mutant mRNA, and blocking translation. Each has the same outcome -- loss of mutant huntingtin.

III. Huntingtin Gene

The disease gene linked to Huntington's disease is termed Huntington or (htt). The huntingtin locus is large, spanning 180 kb and consisting of 67 exons. The huntingtin gene is widely expressed and is required for normal development. It is expressed as 2 alternatively polyadenylated forms displaying different relative abundance in various fetal and adult tissues. The larger transcript is approximately 13.7 kb and is expressed predominantly in adult and fetal brain whereas the smaller transcript of approximately 10.3 kb is more widely expressed. The two transcripts differ with respect to their 3' untranslated regions (Lin et al., 1993). Both messages are predicted to encode a 348 kilodalton protein containing 3144 amino acids. The genetic defect leading to Huntington's disease is believed to confer a new property on the mRNA or alter the function of the protein. The amino acid sequence of the human huntingtin protein is set forth in Figure 2 (SEQ ID NO:2).

A consensus nucleotide sequence of the human huntingtin gene (cDNA) is set forth in Figure 1 (SEQ ID NO:1). The coding region consists of nucleotides 316 to 9750 of SEQ ID NO:1. The two alternative polyadenylation signals are found at nucleotides 10326 to 10331 and nucleotides 13644 to 13649, respectively. The corresponding two polyadenylation sites are found at nucleotides 10348 and 13672, respectively. The first polyadenylation signal/site is that of the 10.3 kb transcript. The second polyadenylation signal/site is that of the 13.7 kb transcript, the predominant transcript in brain.

Five (5) polymorphisms in the human htt gene were identified as described in Example I. An additional 38 polymorphisms in the huntingtin gene sequence have been identified *via* SNP (single nucleotide polymorphism) analysis (see Table 3). The polymorphisms set forth in Tables 2 and 3 represent preferred sites to target *via* single-nucleotide-specific RNAi, as described herein.

Table 2. Polymorphic sites (P) in the htt gene of human cell lines.

Cell line	P1 (2886)	P2 (4034)	P3 (6912)	P4 (7222)	P5 (7246)

GFP-Htt (9kb construct)	C	G	A	T	C
HeLa	t	a	A	g	C
HEK 293T	t	a	G	g	t
HepG2	t	a	G	g	t
FP-4	t	a	g, A	g	t, C

Table 3. Polymorphic sites (P) in the human htt gene identified by SNP analysis.

	consensus	polymorphism	db_xref
complement 103	G	A P6	dbSNP:396875
complement 432	T	C P7	dbSNP:473915
complement 474	C	A P8	dbSNP:603765
1509	T	C P9	dbSNP:1065745
complement 1857	T	C P10	dbSNP:2301367
3565	G	C, A P11,P12	dbSNP:1065746
3594	T	G P13	dbSNP:1143646
3665	G	C P14	dbSNP:1065747
complement 4122	G	A P15	dbSNP:363099
complement 4985	G	A P16	dbSNP:363129
complement 5480	T	G P17	dbSNP:363125
6658	T	G P18	dbSNP:1143648
complement 6912	T	C P19	dbSNP:362336
complement 7753	G	A P20	dbSNP:3025816
complement 7849	G	C P21	dbSNP:3025814
complement 8478	T	C P22	dbSNP:2276881
8574	T	C P23	dbSNP:2229985
complement 9154	C	A P24	dbSNP:3025807
9498	T	C P25	dbSNP:2229987
complement 9699	G	A P26	dbSNP:362308
complement 9809	G	A P27	dbSNP:362307
complement 10064	T	C P28	dbSNP:362306
complement 10112	G	C P29	dbSNP:362268
complement 10124	G	C P30	dbSNP:362305
complement 10236	T	G P31	dbSNP:362304
complement 10271	G	A P32	dbSNP:362303
complement 10879	G	A P33	dbSNP:1557210
complement 10883	G	A P34	dbSNP:362302
complement 10971	C	A P35	dbSNP:3025805
complement 11181	G	A P36	dbSNP:362267
complement 11400	C	A P37	dbSNP:362301
11756..11757	G	- P38	dbSNP:5855774
12658	G	A P39	dbSNP:2237008
complement 12911	T	C P40	dbSNP:362300
complement 13040	G	A P41	dbSNP:2530595
13482	G	A P42	dbSNP:1803770
13563	G	A P43	dbSNP:1803771

The present invention targets mutant huntingtin using RNA interference (Hutvagner et al., 2002). One strand of double-stranded RNA (siRNA) complements a polymorphic region within the mutant huntingtin mRNA. After introduction of siRNA into neurons, the siRNA partially unwinds, binds to polymorphic region within the huntingtin mRNA in a site-specific manner, and activates an mRNA nuclease. This nuclease cleaves the huntingtin mRNA, thereby halting translation of the mutant huntingtin. Cells rid themselves of partially digested mRNA, thus precluding translation, or cells digest partially translated proteins. Neurons survive on the wild-type huntingtin (from the normal allele); this approach prevents the ravages of mutant huntingtin by eliminating its production.

IV. siRNA Design

siRNAs are designed as follows. First, a portion of the target gene (e.g., the htt gene) is selected that includes the polymorphism. Exemplary polymorphisms are selected from the 5' untranslated region of a target gene. Cleavage of mRNA at these sites should eliminate translation of corresponding mutant protein. Polymorphisms from other regions of the mutant gene are also suitable for targeting. A sense strand is designed based on the sequence of the selected portion. Preferably the portion (and corresponding sense strand) includes about 19 to 25 nucleotides, e.g., 19, 20, 21, 22, 23, 24 or 25 nucleotides. More preferably, the portion (and corresponding sense strand) includes 21, 22 or 23 nucleotides. The skilled artisan will appreciate, however, that siRNAs having a length of less than 19 nucleotides or greater than 25 nucleotides can also function to mediate RNAi. Accordingly, siRNAs of such length are also within the scope of the instant invention provided that they retain the ability to mediate RNAi. Longer RNAi agents have been demonstrated to ellicit an interferon or PKR response in certain mammalian cells which may be undesirable. Preferably the RNAi agents of the invention do not ellicit a PKR response (*i.e.*, are of a sufficiently short length). However, longer RNAi agents may be useful, for example, in cell types incapable of generating a PRK response or in situations where the PKR response has been downregulated or dampened by alternative means.

The sense strand sequence is designed such that the polymorphism is essentially in the middle of the strand. For example, if a 21-nucleotide siRNA is chosen, the polymorphism is at, for example, nucleotide 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16 (*i.e.*,

6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16 nucleotides from the 5' end of the sense strand. For a 22-nucleotide siRNA, the polymorphism is at, for example, nucleotide 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16. For a 23-nucleotide siRNA, the polymorphism is at, for example, 7, 8, 9, 10, 11, 12, 13, 14, 15 or 16. For a 24-nucleotide siRNA, the polymorphism is at, for example, 9, 10, 11, 12, 13, 14 or 16. For a 25-nucleotide siRNA, the polymorphism is at, for example, 9, 10, 11, 12, 13, 14, 15, 16 or 17. Moving the polymorphism to an off-center position may, in some instances, reduce efficiency of cleavage by the siRNA. Such compositions, *i.e.*, less efficient compositions, may be desireable for use if off-silencing of the wild-type mRNA is detected.

The antisense strand is routinely the same length as the sense strand and include complementary nucleotides. In one embodiment, the strands are fully complementary, *i.e.*, the strands are blunt-ended when aligned or annealed. In another embodiment, the strands comprise align or anneal such that 1-, 2- or 3-nucleotide overhangs are generated, *i.e.*, the 3' end of the sense strand extends 1, 2 or 3 nucleotides further than the 5' end of the antisense strand and/or the 3' end of the antisense strand extends 1, 2 or 3 nucleotides further than the 5' end of the sense strand. Overhangs can comprise (or consist of) nucleotides corresponding to the target gene sequence (or complement thereof). Alternatively, overhangs can comprise (or consist of) deoxyribonucleotides, for example dTs, or nucleotide analogs, or other suitable non-nucleotide material.

To facilitate entry of the antisense strand into RISC (and thus increase or improve the efficiency of target cleavage and silencing), the base pair strength between the 5' end of the sense strand and 3' end of the antisense strand can be altered, e.g., lessened or reduced, as described in detail in U.S. Provisional patent application nos. 60/475,386 entitled "*Methods and Compositions for Controlling Efficacy of RNA Silencing*" (filed June 2, 2003) and 60/475,331 entitled "*Methods and Compositions for Enhancing the Efficacy and Specificity of RNAi*" (filed June 2, 2003), the contents of which are incorporated in their entirety by this reference. In one embodiment of these aspects of the invention, the base-pair strength is less due to fewer G:C base pairs between the 5' end of the first or antisense strand and the 3' end of the second or sense strand than between the 3' end of the first or antisense strand and the 5' end of the second or sense strand. In another embodiment, the base pair strength is less due to at least one mismatched base pair between the 5' end of the first or antisense strand and the 3' end of the second or sense strand. Preferably, the mismatched base pair is selected

from the group consisting of G:A, C:A, C:U, G:G, A:A, C:C and U:U. In another embodiment, the base pair strength is less due to at least one wobble base pair, e.g., G:U, between the 5' end of the first or antisense strand and the 3' end of the second or sense strand. In another embodiment, the base pair strength is less due to at least one base pair comprising a rare nucleotide, e.g., inosine (I). Preferably, the base pair is selected from the group consisting of an I:A, I:U and I:C. In yet another embodiment, the base pair strength is less due to at least one base pair comprising a modified nucleotide. In preferred embodiments, the modified nucleotide is selected from the group consisting of 2-amino-G, 2-amino-A, 2,6-diamino-G, and 2,6-diamino-A.

The design of siRNAs suitable for targeting the htt polymorphisms set forth in Table 2 is described in detail below

P1 DNA	TGTGCTGAC*T*CTGAGGAACAG	(SEQ ID NO:5)	
sense	UGUGCUGAC*U*CUGAGGAACAG	(SEQ ID NO:6)	
antisense	ACACGACUGAGACUCCUUGUC	(blunt-ends, 21-mer)	(SEQ ID NO:7)

(2-nt overhangs) see Figure 5

P2 DNA	CATACCTCA*A*ACTGCATGATG	(SEQ ID NO:8)	
sense	CAUACCUCA*A*ACUGCAUGAUG	(SEQ ID NO:9)	
antisense	GUAUGGAGUUUGACGUACUAC	(blunt ends, 21-mer)	(SEQ ID NO:10)

P3 DNA	GCCTGCAGA*G*CCGGCGGCCTA	(SEQ ID NO:11)	
sense	GCCUGCAGA*G*CCGGCGGCCUA	(SEQ ID NO:12)	
antisense	CGGACGUCUCGGCCGCCGGAU	(blunt ends, 21-mer)	(SEQ ID NO:13)

P4 DNA	A*C*AGAGTTT*G*TGACCCACGCC	(SEQ ID NO:14)	
sense	ACAGAGUUU*G*UGACCCACGCC	(SEQ ID NO:15)	
antisense	UGUCUCAAACACUGGGUGCGG	(blunt ends, 21-mer)	(SEQ ID NO:16)

P5 DNA	TCCCTCATC*T*ACTGTGTGCAC	(SEQ ID NO:17)	
sense	UCCCUCAUC*U*ACUGUGUGCAC	(SEQ ID NO:18)	
antisense	AGGGAGUAGAUGACACACGUG	(blunt ends, 21 mer)	(SEQ ID NO:19)

siRNAs can be designed according to the above exemplary teachings for any other polymorphisms found in the htt gene. Moreover, the technology is applicable to targeting any other disease gene having associated polymorphisms, *i.e.*, non-disease causing polymorphisms.

To validate the effectiveness by which siRNAs destroy mutant mRNAs (e.g., mutant huntingtin mRNA), the siRNA is incubated with mutant cDNA (e.g., mutant huntingtin cDNA) in a *Drosophila*-based *in vitro* mRNA expression system. Radiolabeled with ^{32}P, newly synthesized mutant mRNAs (e.g., mutant huntingtin mRNA) are detected autoradiographically on an agarose gel. The presence of cleaved mutant mRNA indicates mRNA nuclease activity. Suitable controls include omission of siRNA and use of wild-type huntingtin cDNA. Alternatively, control siRNAs are selected having the same nucleotide composition as the selected siRNA, but without significant sequence complementarity to the appropriate target gene. Such negative controls can be designed by randomly scrambling the nucleotide sequence of the selected siRNA; a homology search can be performed to ensure that the negative control lacks homology to any other gene in the appropriate genome. In addition, negative control siRNAs can be designed by introducing one or more base mismatches into the sequence.

Sites of siRNA-mRNA complementation are selected which result in optimal mRNA specificity and maximal mRNA cleavage.

While the instant invention primarily features targeting polymorphic regions in the target mutant gene (e.g., in mutant htt) distinct from the expanded CAG region mutation, the skilled artisan will appreciate that targeting the mutant region may have applicability as a therapeutic strategy in certain situations. Targeting the mutant region can be accomplished using siRNA that complements CAG in series. The siRNA^{cag} would bind to mRNAs with CAG complementation, but might be expected to have greater opportunity to bind to an extended CAG series. Multiple siRNA^{cag} would bind to the mutant huntingtin mRNA (as opposed to fewer for the wild type huntingtin mRNA); thus, the mutant huntingtin mRNA is more likely to be cleaved. Successful mRNA inactivation using this approach would also eliminate normal or wild-type huntingtin mRNA. Also inactivated, at least to some extent, could be other normal genes (approximately 70) which also have CAG repeats, where their mRNAs could interact with the siRNA. This approach would thus rely on an attrition strategy -- more of the mutant huntingtin mRNA would be destroyed than wild type huntingtin mRNA or the other approximately 69 mRNAs that code for polyglutamines.

V. RNAi Agents

The present invention includes siRNA molecules designed, for example, as described above. The siRNA molecules of the invention can be chemically synthesized, or can be transcribed *in vitro* from a DNA template, or *in vivo* from e.g., shRNA, or, by using recombinant human DICER enzyme, to cleave *in vitro* transcribed dsRNA templates into pools of 20- ,21- or 23- bp duplex RNA mediating RNAi. The siRNA molecules can be designed using any method known in the art.

In one aspect, instead of the RNAi agent being an interfering ribonucleic acid, e.g., an siRNA or shRNA as described above, the RNAi agent can encode an interfering ribonucleic acid, e.g., an shRNA, as described above. In other words, the RNAi agent can be a transcriptional template of the interfering ribonucleic acid. Thus, RNAi agents of the present invention can also include small hairpin RNAs (shRNAs), and expression constructs engineered to express shRNAs. Transcription of shRNAs is initiated at a polymerase III (pol III) promoter, and is thought to be terminated at position 2 of a 4-5-thymine transcription termination site. Upon expression, shRNAs are thought to fold into a stem-loop structure with 3' UU-overhangs; subsequently, the ends of these shRNAs are processed, converting the shRNAs into siRNA-like molecules of about 21-23 nucleotides (Brummelkamp et al., 2002; Lee et al., 2002. *supra*; Miyagishi et al., 2002; Paddison et al., 2002, *supra*; Paul et al., 2002, *supra*; Sui et al., 2002 *supra*; Yu et al., 2002, *supra*. More information about shRNA design and use can be found on the internet at the following addresses: katahdin.cshl.org:9331/RNAi/docs/BseRI-BamHI_Strategy.pdf and katahdin.cshl.org:9331/RNAi/docs/Web_version_of_PCR_strategy1.pdf.

Expression constructs of the present invention include any construct suitable for use in the appropriate expression system and include, but are not limited to, retroviral vectors, linear expression cassettes, plasmids and viral or virally-derived vectors, as known in the art. Such expression constructs can include one or more inducible promoters, RNA Pol III promoter systems such as U6 snRNA promoters or H1 RNA polymerase III promoters, or other promoters known in the art. The constructs can include one or both strands of the siRNA. Expression constructs expressing both strands can also include loop structures linking both strands, or each strand can be separately transcribed from separate promoters within the same construct. Each strand can also be transcribed from a separate expression construct. (Tuschl, T., 2002, *supra*).

Synthetic siRNAs can be delivered into cells by methods known in the art, including cationic liposome transfection and electroporation. However, these exogenous siRNA generally show short term persistence of the silencing effect (4~5 days in cultured cells), which may be beneficial in only certain embodiments. To obtain longer term suppression of the target genes (*i.e.*, mutant genes) and to facilitate delivery under certain circumstances, one or more siRNA can be expressed within cells from recombinant DNA constructs. Such methods for expressing siRNA duplexes within cells from recombinant DNA constructs to allow longer-term target gene suppression in cells are known in the art, including mammalian Pol III promoter systems (e.g., H1 or U6/snRNA promoter systems (Tuschl, T., 2002, *supra*) capable of expressing functional double-stranded siRNAs; (Bagella et al.,1998; Lee et al., 2002, *supra*; Miyagishi et al., 2002, *supra;* Paul et al., 2002, *supra*; Yu et al., 2002), *supra*; Sui et al., 2002, *supra*). Transcriptional termination by RNA Pol III occurs at runs of four consecutive T residues in the DNA template, providing a mechanism to end the siRNA transcript at a specific sequence. The siRNA is complementary to the sequence of the target gene in 5'-3' and 3'-5' orientations, and the two strands of the siRNA can be expressed in the same construct or in separate constructs. Hairpin siRNAs, driven by H1 or U6 snRNA promoter and expressed in cells, can inhibit target gene expression (Bagella et al.,1998; Lee et al., 2002, *supra*; Miyagishi et al., 2002, *supra;* Paul et al., 2002, *supra*; Yu et al., 2002), *supra*; Sui et al., 2002, *supra*). Constructs containing siRNA sequence under the control of T7 promoter also make functional siRNAs when cotransfected into the cells with a vector expressing T7 RNA polymerase (Jacque et al., 2002, *supra*). A single construct may contain multiple sequences coding for siRNAs, such as multiple regions of the gene encoding mutant htt, targeting the same gene or multiple genes, and can be driven, for example, by separate PolIII promoter sites.

Animal cells express a range of noncoding RNAs of approximately 22 nucleotides termed micro RNA (miRNAs) which can regulate gene expression at the post transcriptional or translational level during animal development. One common feature of miRNAs is that they are all excised from an approximately 70 nucleotide precursor RNA stem-loop, probably by Dicer, an RNase III-type enzyme, or a homolog thereof. By substituting the stem sequences of the miRNA precursor with sequence complementary to the target mRNA, a vector construct that expresses the engineered precursor can be used to produce siRNAs to initiate RNAi against specific mRNA

targets in mammalian cells (Zeng et al., 2002, *supra*). When expressed by DNA vectors containing polymerase III promoters, micro-RNA designed hairpins can silence gene expression (McManus et al., 2002, *supra*). MicroRNAs targeting polymorphisms may also be useful for blocking translation of mutant proteins, in the absence of siRNA-mediated gene-silencing. Such applications may be useful in situations, for example, where a designed siRNA caused off-target silencing of wild type protein.

Viral-mediated delivery mechanisms can also be used to induce specific silencing of targeted genes through expression of siRNA, for example, by generating recombinant adenoviruses harboring siRNA under RNA Pol II promoter transcription control (Xia et al., 2002, *supra*). Infection of HeLa cells by these recombinant adenoviruses allows for diminished endogenous target gene expression. Injection of the recombinant adenovirus vectors into transgenic mice expressing the target genes of the siRNA results in *in vivo* reduction of target gene expression. *Id.* In an animal model, whole-embryo electroporation can efficiently deliver synthetic siRNA into post-implantation mouse embryos (Calegari et al., 2002). In adult mice, efficient delivery of siRNA can be accomplished by "high-pressure" delivery technique, a rapid injection (within 5 seconds) of a large volume of siRNA containing solution into animal *via* the tail vein (Liu et al.,1999, *supra;* McCaffrey et al., 2002, *supra*; Lewis et al., 2002. Nanoparticles and liposomes can also be used to deliver siRNA into animals.

The nucleic acid compositions of the invention include both unmodified siRNAs and modified siRNAs as known in the art, such as crosslinked siRNA derivatives or derivatives having non nucleotide moieties linked, for example to their 3' or 5' ends. Modifying siRNA derivatives in this way may improve cellular uptake or enhance cellular targeting activities of the resulting siRNA derivative as compared to the corresponding siRNA, are useful for tracing the siRNA derivative in the cell, or improve the stability of the siRNA derivative compared to the corresponding siRNA.

Engineered RNA precursors, introduced into cells or whole organisms as described herein, will lead to the production of a desired siRNA molecule. Such an siRNA molecule will then associate with endogenous protein components of the RNAi pathway to bind to and target a specific mRNA sequence for cleavage and destruction. In this fashion, the mRNA to be targeted by the siRNA generated from the engineered RNA precursor will be depleted from the cell or organism, leading to a decrease in the concentration of the protein encoded by that mRNA in the cell or organism. The RNA

precursors are typically nucleic acid molecules that individually encode either one strand of a dsRNA or encode the entire nucleotide sequence of an RNA hairpin loop structure.

The nucleic acid compositions of the invention can be unconjugated or can be conjugated to another moiety, such as a nanoparticle, to enhance a property of the compositions, e.g., a pharmacokinetic parameter such as absorption, efficacy, bioavailability, and/or half-life. The conjugation can be accomplished by methods known in the art, e.g., using the methods of Lambert *et al.*, Drug Deliv. Rev.:47(1), 99-112 (2001) (describes nucleic acids loaded to polyalkylcyanoacrylate (PACA) nanoparticles); Fattal *et al.*, J. Control Release 53(1-3):137-43 (1998) (describes nucleic acids bound to nanoparticles); Schwab *et al.*, Ann. Oncol. 5 Suppl. 4:55-8 (1994) (describes nucleic acids linked to intercalating agents, hydrophobic groups, polycations or PACA nanoparticles); and Godard *et al.*, Eur. J. Biochem. 232(2):404-10 (1995) (describes nucleic acids linked to nanoparticles).

The nucleic acid molecules of the present invention can also be labeled using any method known in the art; for instance, the nucleic acid compositions can be labeled with a fluorophore, e.g., Cy3, fluorescein, or rhodamine. The labeling can be carried out using a kit, e.g., the SILENCER™ siRNA labeling kit (Ambion). Additionally, the siRNA can be radiolabeled, e.g., using ^{3}H, ^{32}P, or other appropriate isotope.

Moreover, because RNAi is believed to progress *via* at least one single-stranded RNA intermediate, the skilled artisan will appreciate that ss-siRNAs (e.g., the antisense strand of a ds-siRNA) can also be designed (e.g., for chemical synthesis) generated (e.g., enzymatically generated)or expressed (e.g., from a vector or plasmid) as described herein and utilized according to the claimed methodologies. Moreover, in invertebrates, RNAi can be triggered effectively by long dsRNAs (e.g., dsRNAs about 100 – 1000 nucleotides in length, preferably about 200- 500, for example, about 250, 300, 350, 400 or 450 nucleotides in length) acting as effectors of RNAi. (Brondani et al., Proc Natl Acad Sci U S A. 2001 Dec 4;98(25):14428-33. Epub 2001 Nov 27).

VI. Methods of Introducing RNAs, Vectors, and Host Cells

Physical methods of introducing nucleic acids include injection of a solution containing the RNA, bombardment by particles covered by the RNA, soaking the cell or organism in a solution of the RNA, or electroporation of cell membranes in the presence of the RNA. A viral construct packaged into a viral particle would accomplish both

efficient introduction of an expression construct into the cell and transcription of RNA encoded by the expression construct. Other methods known in the art for introducing nucleic acids to cells may be used, such as lipid-mediated carrier transport, chemical-mediated transport, such as calcium phosphate, and the like. Thus the RNA may be introduced along with components that perform one or more of the following activities: enhance RNA uptake by the cell, inhibit annealing of single strands, stabilize the single strands, or other-wise increase inhibition of the target gene.

RNA may be directly introduced into the cell (*i.e.*, intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism, introduced orally, or may be introduced by bathing a cell or organism in a solution containing the RNA. Vascular or extravascular circulation, the blood or lymph system, and the cerebrospinal fluid are sites where the RNA may be introduced.

The cell having the target gene may be from the germ line or somatic, totipotent or pluripotent, dividing or non-dividing, parenchyma or epithelium, immortalized or transformed, or the like. The cell may be a stem cell or a differentiated cell. Cell types that are differentiated include adipocytes, fibroblasts, myocytes, cardiomyocytes, endothelium, neurons, glia, blood cells, megakaryocytes, lymphocytes, macrophages, neutrophils, eosinophils, basophils, mast cells, leukocytes, granulocytes, keratinocytes, chondrocytes, osteoblasts, osteoclasts, hepatocytes, and cells of the endocrine or exocrine glands.

Depending on the particular target gene and the dose of double stranded RNA material delivered, this process may provide partial or complete loss of function for the target gene. A reduction or loss of gene expression in at least 50%, 60%, 70%, 80%, 90%, 95% or 99% or more of targeted cells is exemplary. Inhibition of gene expression refers to the absence (or observable decrease) in the level of protein and/or mRNA product from a target gene. Specificity refers to the ability to inhibit the target gene without manifest effects on other genes of the cell. The consequences of inhibition can be confirmed by examination of the outward properties of the cell or organism (as presented below in the examples) or by biochemical techniques such as RNA solution hybridization, nuclease protection, Northern hybridization, reverse transcription, gene expression monitoring with a microarray, antibody binding, enzyme linked immunosorbent assay (ELISA), Western blotting, radioimmunoassay (RIA), other immunoassays, and fluorescence activated cell analysis (FACS).

For RNA-mediated inhibition in a cell line or whole organism, gene expression is conveniently assayed by use of a reporter or drug resistance gene whose protein product is easily assayed. Such reporter genes include acetohydroxyacid synthase (AHAS), alkaline phosphatase (AP), beta galactosidase (LacZ), beta glucoronidase (GUS), chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), horseradish peroxidase (HRP), luciferase (Luc), nopaline synthase (NOS), octopine synthase (OCS), and derivatives thereof. Multiple selectable markers are available that confer resistance to ampicillin, bleomycin, chloramphenicol, gentamycin, hygromycin, kanamycin, lincomycin, methotrexate, phosphinothricin, puromycin, and tetracyclin. Depending on the assay, quantitation of the amount of gene expression allows one to determine a degree of inhibition which is greater than 10%, 33%, 50%, 90%, 95% or 99% as compared to a cell not treated according to the present invention. Lower doses of injected material and longer times after administration of RNAi agent may result in inhibition in a smaller fraction of cells (e.g., at least 10%, 20%, 50%, 75%, 90%, or 95% of targeted cells). Quantization of gene expression in a cell may show similar amounts of inhibition at the level of accumulation of target mRNA or translation of target protein. As an example, the efficiency of inhibition may be determined by assessing the amount of gene product in the cell; mRNA may be detected with a hybridization probe having a nucleotide sequence outside the region used for the inhibitory double-stranded RNA, or translated polypeptide may be detected with an antibody raised against the polypeptide sequence of that region.

The RNA may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses (e.g., at least 5, 10, 100, 500 or 1000 copies per cell) of material may yield more effective inhibition; lower doses may also be useful for specific applications.

In a preferred aspect, the efficacy of an RNAi agent of the invention (e.g., an siRNA targeting a polymorphism in a mutant gene) is tested for its ability to specifically degrade mutant mRNA (e.g., mutant htt mRNA and/or the production of mutant huntingtin protein) in cells, in particular, in neurons (e.g., striatal or cortical neuronal clonal lines and/or primary neurons). Also suitable for cell-based validation assays are other readily transfectable cells, for example, HeLa cells or COS cells. Cells are transfected with human wild type or mutant cDNAs (e.g., human wild type or mutant huntingtin cDNA). Standard siRNA, modified siRNA or vectors able to produce siRNA

from U-looped mRNA are co-transfected. Selective reduction in mutant mRNA (e.g., mutant huntingtin mRNA) and/or mutant protein (e.g., mutant huntingtin) is measured. Reduction of mutant mRNA or protein can be compared to levels of normal mRNA or protein. Exogenously-introduced normal mRNA or protein (or endogenous normal mRNA or protein) can be assayed for comparison purposes. When utilizing neuronal cells, which are known to be somewhat resistant to standard transfection techniques, it may be desirable to introduce RNAi agents (e.g., siRNAs) by passive uptake.

VII. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disease or disorder caused, in whole or in part, by a gain of function mutant protein. In one embodiment, the disease or disorder is a trinucleotide repeat disease or disorder. In another embodiment, the disease or disorder is a polyglutamine disorder. In a preferred embodiment, the disease or disorder is a disorder associated with the expression of huntingtin and in which alteration of huntingtin, especially the amplification of CAG repeat copy number, leads to a defect in huntingtin gene (structure or function) or huntingtin protein (structure or function or expression), such that clinical manifestations include those seen in Huntington's disease patients.

"Treatment", or "treating" as used herein, is defined as the application or administration of a therapeutic agent (e.g., a RNA agent or vector or transgene encoding same) to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has the disease or disorder, a symptom of disease or disorder or a predisposition toward a disease or disorder, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease or disorder, the symptoms of the disease or disorder, or the predisposition toward disease.

In one aspect, the invention provides a method for preventing in a subject, a disease or disorder as described above, by administering to the subject a therapeutic agent (e.g., an RNAi agent or vector or transgene encoding same). Subjects at risk for the disease can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the disease or disorder, such that the disease or disorder is prevented or, alternatively, delayed in its progression.

Another aspect of the invention pertains to methods treating subjects therapeutically, *i.e.,* alter onset of symptoms of the disease or disorder. In an exemplary embodiment, the modulatory method of the invention involves contacting a cell expressing a gain-of-function mutant with a therapeutic agent (e.g., a RNAi agent or vector or transgene encoding same) that is specific for a polymorphism within the gene, such that sequence specific interference with the gene is achieved. These methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the target gene molecules of the present invention or target gene modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

Therapeutic agents can be tested in an appropriate animal model. For example, an RNAi agent (or expression vector or transgene encoding same) as described herein can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with said agent. Alternatively, a therapeutic agent can be used in an animal model to determine the mechanism of action of such an agent. For example, an agent can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent can be used in an animal model to determine the mechanism of action of such an agent.

VIII. Pharmaceutical Compositions

The invention pertains to uses of the above-described agents for prophylactic and/or therapeutic treatments as described infra. Accordingly, the modulators (e.g., RNAi agents) of the present invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, antibody, or modulatory compound and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, intraperitoneal, intramuscular, oral (e.g., inhalation), transdermal (topical), and transmucosal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all

cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as

microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject

to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds that exhibit large therapeutic indices are preferred. Although compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the EC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal response) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

Unlike other types of autosomal dominant diseases, Huntington's disease does not contain a point mutation e.g.,) single nucleotide change. Therefore, the strategy to design siRNA directed against a point mutation in the disease allele cannot be implemented. Instead, the present invention directs designed siRNAs against polymorphisms in the Huntingtin gene, of which there are about 30 available in GenBank. The present invention also identifies the polymorphism in the Huntington disease allele which differs from the wild type allele, so that siRNA destroys only the disease mRNA and leaves intact the wild type (normal) allele mRNA. Thus, only the mutant Huntingtin protein is destroyed and the normal protein is intact.

Example I: Testing of RNAi agents (e.g., siRNAs) against mutant htt in *Drosophila* lysates

A siRNA targeting position 2886 in the htt mRNA was designed as described *supra*. The sequence of the siRNA is depicted in Figure 5a (SEQ ID NO:24 sense; 25 anti-sense). Synthetic RNA (Dharmacon) was deprotected according to the manufacturer's protocol. siRNA strands were annealed (Elbashir et al., 2001a).

Target RNAs were prepared as follows. Target RNAs were transcribed with recombinant, histidine-tagged, T7 RNA polymerase from PCR products as described (Nykänen et al., 2001; Hutvágner et al., 2002). PCR templates for htt sense and anti-sense were generated by amplifying 0.1 ng/ml (final concentration) plasmid template encoding htt cDNA using the following primer pairs: htt sense target, 5′-GCG TAA TAC GAC TCA CTA TAG GAA CAG TAT GTCTCA GAC ATC-3′ (SEQ ID NO:30) and 5′-UUCG AAG UAU UCC GCG UAC GU-3′ (SEQ ID NO:31); htt anti-sense target, 5′-GCG TAA TAC GAC TCA CTA TAG GAC AAG CCT AAT TAG TGA TGC-3′ (SEQ ID NO:32).and 5′-GAA CAG TAT GTC TCA GAC ATC-3′ (SEQ ID NO:33).

The siRNA was tested using an *in vitro* RNAi assay, featuring *Drosophila* embryo lysates. *In vitro* RNAi reactions and analysis was carried out as previously described (Tuschl et al., 1999; Zamore et al., 2000; Haley et al., 2003). Target RNAs were used at ~ 5 nM concentration so that reactions are mainly under single-turnover conditions. Target cleavage under these conditions is proportionate to siRNA concentration.

Figure 5a shows the efficacy of the siRNA directed against position 2886 in the mutant htt. The data clearly demonstrate that the siRNA directs cleavage of the sense target to a greater degree than observed for the anti-sense target. However, it is noticed that this first-designed siRNA did not produce a very active molecule, at least in this *in vitro* assay. Thermodynamic analysis of the base pair strength at the two ends of the siRNA duplex indicated roughly equivalent base pair strengths. Figure 4 depicts the thermodynamic analysis of siRNA sense (SEQ ID NO:20; 22 respectively) and anti-sense (SEQ ID NO:21; 23 respectively) strand 5' ends for the siRNA duplex in 5a. ΔG (kcal/mole) was calculated in 1M NaCl at 37°C.

To improved the efficacy of the designed siRNA duplex, the 5′ end of the sense strand or position 19 of the anti-sense strand of the htt siRNA tested in Figure 5a was altered to produce siRNA duplexes in which the 5′ end of the sense strand was either fully unpaired (Figure 5c; SEQ ID NO: 28 sense; SEQ ID NO:29 anti-sense) or in an A:U base pair (Figure 5b; SEQ ID NO:26 sense; SEQ ID NO:27 anti-sense). The unpairing the 5′ end of an siRNA strand-the sense strand, in this case-causes that strand to function to the exclusion of the other strand. When the htt sense strand 5′ end was present in an A:U base pair and the htt anti-sense strand 5′ end was in a G:C pair, the sense strand dominated the reaction (Figure 5b-c), but the htt anti-sense strand retained activity similar to that seen for the originally-designed siRNA.

Example II: RNAi knockdown of Htt protein in cultured cells

In a first experiment, siRNAs targeting a polymorphism in the htt mRNA (*i.e.*, the polymorphism at position 2886 in the htt mRNA) were tested for their ability to down-regulate endogenous Htt protein in HeLa cells. HeLa cells were cultures and transfected as follows. HeLa cells were maintained at 37°C in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10% fetal bovine serum (FBS), 100 unit/ml penicillin and 100 μg/ml streptomycin (Invitrogen). Cells were

regularly passaged at sub-confluence and plated at 70% confluency 16 hours before transfection. Lipofectamine™ (Invitrogen)-mediated transient transfection of siRNAs were performed in duplicate 6-well plates (Falcon) as described for adherent cell lines by the manufacturer. A standard transfection mixture containing 100-150 nM siRNA and 9-10 µl Lipofectamine™ in 1 ml serum-reduced OPTI-MEM® (Invitrogen) was added to each well. Cells were incubated in transfection mixture at 37C for 6 hours and further cultured in antibiotic-free DMEM. For Western blot analysis at various time intervals, the transfected cells were harvested, washed twice with phosphate buffered saline (PBS, Invitrogen), flash frozen in liquid nitrogen, and stored at -80°C for analysis.

Three siRNAs were tested against a common target sequence in exon 1 and four siRNAs were tested for the position 2886 polymorphism. Western blot analysis was performed as follows. Cells treated with siRNA were harvested as described above and lysed in ice-cold reporter lysis buffer (Promega) containing protease inhibitor (complete, EDTA-free, 1 tablet/10 ml buffer, Roche Molecular Biochemicals). After clearing the resulting lysates by centrifugation, protein in clear lysates was quantified by Dc protein assay kit (Bio-Rad). Proteins in 60 µg of total cell lysate were resolved by 10% SDS-PAGE, transferred onto a polyvinylidene difluoride membrane (PVDF, Bio-Rad), and immuno-blotted with antibodies against CD80 (Santa Cruz). Protein content was visualized with a BM Chemiluminescence Blotting Kit (Roche Molecular Biochemicals). The blots were exposed to x-ray film (Kodak MR-1) for various times (30 s to 5 min). Figure 6a depicts the results of the Western analysis. Tubulin served as the loading control. The data are quantified and normalized in Figure 6b. Of the siRNAs tested, 2886-4, reproducibly showed enhanced efficacy in cultured HeLa cells (Figure 6). This siRNA also reproducibly showed enhanced efficacy *in vitro* (not shown). GFP siRNA is a control siRNA that shares no sequence homology with htt mRNA.

siRNAs against polymorphic regions in the htt mRNA can likewise be tested in cells transfected with human htt cDNA or in cells transfected with htt reporter constructs. Lipofectamine™ (Invitrogen)-mediated transient cotransfections of cDNAs or reporter plasmids and siRNAs are performed as described *supra*. To test the ability of siRNAs to target htt reported constructs, RNAi was used to inhibit GFP-htt expression in cultured human Hela cell lines. Briefly, HeLa cells were transfected with GFP-htt siRNA duplex, targeting the GFP-htt mRNA sequence. To analyze RNAi effects against

GFP-htt, lysates were prepared from siRNA duplex-treated cells at various times after transfection. Western blot experiments were carried out as described supra. Briefly, HeLa cells were harvested at various times post transfection, their protein content was resolved on 10% SDS-PAGE, transferred onto PVDF membranes, and immunoblotted with appropriate antibodies. Results of this study indicated that siRNA against GFP can eliminate expression of GFP-htt expression in Hela cells transfected with the GFP-htt gene. For studies targeting exogenously introduces htt, procedures are as described except that anti-Htt antibodies are used for immunoblotting.

RNAi can be used to inhibit htt expression in cultured neuronal cells as well. Exemplary cells include PC12 (Scheitzer et al., Thompson et al.) and NT3293 (Tagle et al.) cell lines as previously described. Additional exemplary cells include stably-transfected cells, *e.g.* neuronal cells or neuronally-derived cells. PC12 cell lines expressing exon 1 of the human huntingtin gene (Htt) can be used although expression of exon 1 reduces cell survival. GFP-Htt PC12 cells having an inducible GFP-Htt gene can also be used to test or validate siRNA efficacy.

Example III: Htt siRNA delivery in an *in vivo* setting

R6/2 mice models (expressing the R6/2 human htt cDNA product) are an accepted animal model to study the effectiveness of siRNA delivery in an *in vivo* setting. Genetically engineered R6/2 mice were used to test the effectiveness of siRNA at the 5' terminus of huntingtin mRNA. Htt siRNA was injected into the striatum of R6/2 mice through an Alzet pump. Mice were treated for 14 days with the siRNA/Alzet pump delivery system.

Results of this study indicated that two mice receiving the siRNA with Trans-IT TKO (Mirus) as either a 20 or 200 nM solution at 0.25µl/hour showed no deterioration of motor impairment from day 67 to day 74. Generally, these R6/2 are expected to have a continued reduction in rotarod beyond day 60.

References

Aronin et al., Neuron, Nov; 15(5):1193-201 (1995)

Aronin et al., Phil Trans Royal Society, June 29; 354 (1386):995-1003 (1999)

Bagella et al., J. Cell. Physiol. 177:206–213 (1998)

Brummelkamp et al., Science 296:550-553 (2002)

Calegari et al., Proc. Natl. Acad. Sci. USA 99(22):14236-40 (2002)

Difiglia et al., Science, Sep 26;277(5334):1990-3 (1997)

Elbashir et al., Genes Dev 15, 188-200 (2001a)

Haley et al., Methods 30, 330-336 (2003)

Hutvágner and Zamore, Science 297, 2056-2060 (2002)

Jacque et al., (2002)

Kremer *et al.*, (1994)

Laforet et al., J. Neurosci., Dec 1;21(23):9112-23 (2001)

Lee et al., EMBO J. 21: 4663-4670.(2002)

Lewis et al., Nature Genetics 32:107-108 (2002)

Lin et al., (1993)

Liu et al., (1999)

McCaffrey et al., Gene Ther. 2002 Dec;9(23):1563 (2002)

McManus et al., RNA 8, 842-850 (2002)

Miyagishi et al.,Nature Biotechnol. 20:497-500 (2002)

Nykänen et al., Cell 107, 309-321 (2001)

Paddison et al., Genes Dev 16, 948-958. (2002)

Paul et al., Nat Biotechnol 20, 505-508 (2002)

Scheitzer et al.

Sui et al., Proc Natl Acad Sci USA 99, 5515-5520 (2002)

Tagle et al.

Thompson et al.

Tuschl, T., Nat Biotechnol. 2002 May;20(5):446-8 (2002)

Tuschl et al., Genes Dev 13, 3191-3197 (1999)

Xia et al., (2002)

Yohrling G.J. et al., Mol Cell Neurosci. May;23(1):28-38 (2003)

Yu et al., Proc Natl Acad Sci U S A 99, 6047-6052 (2002)

Zamore et al., Cell 101, 25-33 (2000)

Zamore et al., Nature Medicine, volume 9 Number 3 pp 266 – 267 (2003)

Zeng et al., (2002)

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

WHAT IS CLAIMED IS:

1. A method of treating a subject having or at risk for a disease characterized or caused by a gain-of-function mutant protein, comprising: administering to said subject an effective amount of a RNAi agent targeting an allelic polymorphism within a gene encoding said mutant protein, such that sequence-specific interference of said gene occurs; thereby treating said disease in said subject.

2. The method of claim 1, wherein said gene comprises an expanded trinucleotide repeat region.

3. The method of claim 1, wherein said mutant protein comprises an expanded polyglutamine domain.

4. The method of claim 1, wherein the disease is selected from the group consisting of Huntington's disease, spino-cerebellar ataxia type 1, spino-cerebellar ataxia type 2, spino-cerebellar ataxia type 3, spino-cerebellar ataxia type 6, spino-cerebellar ataxia type 7, spino-cerebellar ataxia type 8, spino-cerebellar ataxia type 12, fragile X syndrome, fragile XE MR, Friedreich ataxia, myotonic dystrophy, spinal bulbar muscular disease and dentatoiubral-pallidoluysian atrophy.

5. The method of claim 4, wherein the disease is Huntington's disease.

6. The method of claim 5, wherein the RNAi agent targets an allelic polymorphism within the gene encoding a huntingtin protein.

7. The method of claim 5, wherein the RNAi agent targets a polymorphism selected from the group consisting of P1-P5.

8. The method of claim 5, wherein the RNAi agent targets a polymorphism selected from the group consisting of P6-P43.

9. The method of claim 1, wherein the RNAi agent comprises a first strand comprising about 16-25 nucleotides homologous to a region of the gene comprising the polymorphism and a second strand comprising about 16-25 nucleotides complementary to the first strand.

10. The method of claim 1, wherein the effective amount is an amount effective to inhibit the expression or activity of the mutant protein.

11. An RNAi agent comprising a first strand comprising about 16-25 nucleotides homologous to a region of a gene encoding a gain-of-function mutant protein, said region comprising an allelic polymorphism, and a second strand comprising about 16-25 nucleotides complementary to the first strand, wherein the RNAi agent direct target-specific cleavage of a mRNA transcribed from the gene encoding the mutant protein.

12. The RNAi agent of claim 9, which targets a polymorphism within the gene encoding a Huntington protein.

13. The RNAi agent of claim 10, wherein said polymorphism is selected from the group comprising P1-P5.

14. The RNAi agent of claim 10, wherein said polymorphism is selected from the group comprising P6-P43.

15. The RNAi agent of any one of claims 11-14, wherein the first strand comprises a nucleotide sequence identical to the sequence of the polymorphism.

16. The RNAi agent of any one of claims 11-14, further comprising a loop portion comprising 4-11 nucleotides that connects the two strands.

17. An isolated nucleic acid molecule encoding the RNAi agent of any one of claims 11-16.

18. A vector comprising the nucleic acid molecule of claim 17.

19. The vector of claim 19, which is a viral vector, retroviral vector, expression cassette, or plasmid.

20. The vector of claim 18, further comprising an RNA Polymerase III or RNA Polymerase II promoter.

21. The vector of claim 18, wherein the RNA Polymerase III promoter is the U6 or H1 promoter.

22. A host cell comprising the RNAi agent or nucleic acid molecule of any one of claims 11-17.

23. A host cell comprising the vector of any one of claims 18-22.

24. The host cell of claim 22 or 23, which is a mammalian host cell.

25. The host cell of claim 24, which is a non-human mammalian cell.

26. The host cell of claim 24, which is a human cell.

27. A composition comprising the RNAi agent or nucleic acid molecule of any one of claims 11-17, and a pharmaceutically acceptable carrier.

28. A method for treating a disease or disorder in a subject caused by a gain-of function mutant protein, comprising identifying an allelic polymorphism within a gene encoding said mutant protein and administering to said subject an RNAi agent targeting said polymorphism such that the mutant protein is decreased, thereby treating the subject.

```
   1 TTGCTGTGTG AGGCAGAACC TGCGGGGGCA GGGGCGGGCT GGTTCCCTGG CCAGCCATTG
  61 GCAGAGTCCG CAGGCTAGGG CTGTCAATCA TGCTGGCCGG CGTGGCCCCG CCTCCGCCGG
 121 CGCGGCCCCG CCTCCGCCGG CGCACGTCTG GGACGCAAGG CGCCGTGGGG GCTGCCGGGA
 181 CGGGTCCAAG ATGGACGGCC GCTCAGGTTC TGCTTTTACC TGCGGCCCAG AGCCCCATTC
 241 ATTGCCCCGG TGCTGAGCGG CGCCGCGAGT CGGCCCGAGG CCTCCGGGGA CTGCCGTGCC
 301 GGGCGGGAGA CCGCCNNNGC GACCCTGGAA AAGCTGATGA AGGCCTTCGA GTCCCTCAAG
 361 TCCTTCCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG
 421 CAGCAGCAGC AACAGCCGCC ACCGCCGCCG CCGCCGCCGC CGCCTCCTCA GCTTCCTCAG
 481 CCGCCGCCGC AGGCACAGCC GCTGCTGCCT CAGCCGCAGC CGCCCCCGCC GCCGCCCCCG
 541 CCGCCACCCG GCCCGGCTGT GGCTGAGGAG CCGCTGCACC GACCAAAGAA AGAACTTTCA
 601 GCTACCAAGA AAGACCGTGT GAATCATTGT CTGACAATAT GTGAAAACAT AGTGGCACAG
 661 TCTGTCAGAA ATTCTCCAGA ATTTCAGAAA CTTCTGGGCA TCGCTATGGA ACTTTTTCTG
 721 CTGTGCAGTG ATGACGCAGA GTCAGATGTC AGGATGGTGG CTGACGAATG CCTCAACAAA
 781 GTTATCAAAG CTTTGATGGA TTCTAATCTT CCAAGGTTAC AGCTCGAGCT CTATAAGGAA
 841 ATTAAAAAGA ATGGTGCCCC TCGGAGTTTG CGTGCTGCCC TGTGGAGGTT TGCTGAGCTG
 901 GCTCACCTGG TTCGGCCTCA GAAATGCAGG CCTTACCTGG TGAACCTTCT GCCGTGCCTG
 961 ACTCGAACAA GCAAGAGACC CGAAGAATCA GTCCAGGAGA CCTTGGCTGC AGCTGTTCCC
1021 AAAATTATGG CTTCTTTTGG CAATTTTGCA AATGACAATG AAATTAAGGT TTTGTTAAAG
1081 GCCTTCATAG CGAACCTGAA GTCAAGCTCC CCCACCATTC GGCGGACAGC GGCTGGATCA
1141 GCAGTGAGCA TCTGCCAGCA CTCAAGAAGG ACACAATATT TCTATAGTTG GCTACTAAAT
1201 GTGCTCTTAG GCTTACTCGT TCCTGTCGAG GATGAACACT CCACTCTGCT GATTCTTGGC
```

FIG.1A

```
1261 GTGCTGCTCA CCCTGAGGTA TTTGGTGCCC TTGCTGCAGC AGCAGGTCAA GGACACAAGC
1321 CTGAAAGGCA GCTTCGGAGT GACAAGGAAA GAAATGGAAG TCTCTCCTTC TGCAGAGCAG
1381 CTTGTCCAGG TTTATGAACT GACGTTACAT CATACACAGC ACCAAGACCA CAATGTTGTG
1441 ACCGGAGCCC TGGAGCTGTT GCAGCAGCTC TTCAGAACGC CTCCACCCGA GCTTCTGCAA
1501 ACCCTGACCG CAGTCGGGGG CATTGGGCAG CTCACCGCTG CTAAGGAGGA GTCTGGTGGC
1561 CGAAGCCGTA GTGGGAGTAT TGTGGAACTT ATAGCTGGAG GGGGTTCCTC ATGCAGCCCT
1621 GTCCTTTCAA GAAAACAAAA AGGCAAAGTG CTCTTAGGAG AAGAAGAAGC CTTGGAGGAT
1681 GACTCTGAAT CGAGATCGGA TGTCAGCAGC TCTGCCTTAA CAGCCTCAGT GAAGGATGAG
1741 ATCAGTGGAG AGCTGGCTGC TTCTTCAGGG GTTTCCACTC CAGGGTCAGC AGGTCATGAC
1801 ATCATCACAG AACAGCCACG GTCACAGCAC ACACTGCAGG CGGACTCAGT GGATCTGGCC
1861 AGCTGTGACT TGACAAGCTC TGCCACTGAT GGGGATGAGG AGGATATCTT GAGCCACAGC
1921 TCCAGCCAGG TCAGCGCCGT CCCATCTGAC CCTGCCATGG ACCTGAATGA TGGGACCCAG
1981 GCCTCGTCGC CCATCAGCGA CAGCTCCCAG ACCACCACCG AAGGGCCTGA TTCAGCTGTT
2041 ACCCCTTCAG ACAGTTCTGA AATTGTGTTA GACGGTACCG ACAACCAGTA TTTGGGCCTG
2101 CAGATTGGAC AGCCCCAGGA TGAAGATGAG GAAGCCACAG GTATTCTTCC TGATGAAGCC
2161 TCGGAGGCCT TCAGGAACTC TTCCATGGCC CTTCAACAGG CACATTTATT GAAAAACATG
2221 AGTCACTGCA GGCAGCCTTC TGACAGCAGT GTTGATAAAT TTGTGTTGAG AGATGAAGCT
2281 ACTGAACCGG GTGATCAAGA AAACAAGCCT TGCCGCATCA AAGGTGACAT TGGACAGTCC
2341 ACTGATGATG ACTCTGCACC TCTTGTCCAT TGTGTCCGCC TTTTATCTGC TTCGTTTTTG
2401 CTAACAGGGG GAAAAAATGT GCTGGTTCCG GACAGGGATG TGAGGGTCAG CGTGAAGGCC
2461 CTGGCCCTCA GCTGTGTGGG AGCAGCTGTG GCCCTCCACC CGGAATCTTT CTTCAGCAAA
```

FIG.1B

```
2521 CTCTATAAAG TTCCTCTTGA CACCACGGAA TACCCTGAGG AACAGTATGT CTCAGACATC
2581 TTGAACTACA TCGATCATGG AGACCCACAG GTTCGAGGAG CCACTGCCAT TCTCTGTGGG
2641 ACCCTCATCT GCTCCATCCT CAGCAGGTCC CGCTTCCACG TGGGAGATTG GATGGGCACC
2701 ATTAGAACCC TCACAGGAAA TACATTTTCT TTGGCGGATT GCATTCCTTT GCTGCGGAAA
2761 ACACTGAAGG ATGAGTCTTC TGTTACTTGC AAGTTAGCTT GTACAGCTGT GAGGAACTGT
2821 GTCATGAGTC TCTGCAGCAG CAGCTACAGT GAGTTAGGAC TGCAGCTGAT CATCGATGTG
2881 CTGACTCTGA GGAACAGTTC CTATTGGCTG GTGAGGACAG AGCTTCTGGA AACCCTTGCA
2941 GAGATTGACT TCAGGCTGGT GAGCTTTTTG GAGGCAAAAG CAGAAAACTT ACACAGAGGG
3001 GCTCATCATT ATACAGGGCT TTTAAAACTG CAAGAACGAG TGCTCAATAA TGTTGTCATC
3061 CATTTGCTTG GAGATGAAGA CCCCAGGGTG CGACATGTTG CCGCAGCATC ACTAATTAGG
3121 CTTGTCCCAA AGCTGTTTTA TAAATGTGAC CAAGGACAAG CTGATCCAGT AGTGGCCGTG
3181 GCAAGAGATC AAAGCAGTGT TTACCTGAAA CTTCTCATGC ATGAGACGCA GCCTCCATCT
3241 CATTTCTCCG TCAGCACAAT AACCAGAATA TATAGAGGCT ATAACCTACT ACCAAGCATA
3301 ACAGACGTCA CTATGGAAAA TAACCTTTCA AGAGTTATTG CAGCAGTTTC TCATGAACTA
3361 ATCACATCAA CCACCAGAGC ACTCACATTT GGATGCTGTG AAGCTTTGTG TCTTCTTTCC
3421 ACTGCCTTCC CAGTTTGCAT TTGGAGTTTA GGTTGGCACT GTGGAGTGCC TCCACTGAGT
3481 GCCTCAGATG AGTCTAGGAA GAGCTGTACC GTTGGGATGG CCACAATGAT TCTGACCCTG
3541 CTCTCGTCAG CTTGGTTCCC ATTGGATCTC TCAGCCCATC AAGATGCTTT GATTTTGGCC
3601 GGAAACTTGC TTGCAGCCAG TGCTCCCAAA TCTCTGAGAA GTTCATGGGC CTCTGAAGAA
3661 GAAGCCAACC CAGCAGCCAC CAAGCAAGAG GAGGTCTGGC CAGCCCTGGG GGACCGGGCC
3721 CTGGTGCCCA TGGTGGAGCA GCTCTTCTCT CACCTGCTGA AGGTGATTAA CATTTGTGCC
```

FIG.1C

```
3781 CACGTCCTGG ATGACGTGGC TCCTGGACCC GCAATAAAGG CAGCCTTGCC TTCTCTAACA
3841 AACCCCCCTT CTCTAAGTCC CATCCGACGA AAGGGGAAGG AGAAAGAACC AGGAGAACAA
3901 GCATCTGTAC CGTTGAGTCC CAAGAAAGGC AGTGAGGCCA GTGCAGCTTC TAGACAATCT
3961 GATACCTCAG GTCCTGTTAC AACAAGTAAA TCCTCATCAC TGGGGAGTTT CTATCATCTT
4021 CCTTCATACC TCAAACTGCA TGATGTCCTG AAAGCTACAC ACGCTAACTA CAAGGTCACG
4081 CTGGATCTTC AGAACAGCAC GGAAAAGTTT GGAGGGTTTC TCCGCTCAGC CTTGGATGTT
4141 CTTTCTCAGA TACTAGAGCT GGCCACACTG CAGGACATTG GGAAGTGTGT TGAAGAGATC
4201 CTAGGATACC TGAAATCCTG CTTTAGTCGA GAACCAATGA TGGCAACTGT TTGTGTTCAA
4261 CAATTGTTGA AGACTCTCTT TGGCACAAAC TTGGCCTCCC AGTTTGATGG CTTATCTTCC
4321 AACCCCAGCA AGTCACAAGG CCGAGCACAG CGCCTTGGCT CCTCCAGTGT GAGGCCAGGC
4381 TTGTACCACT ACTGCTTCAT GGCCCCGTAC ACCCACTTCA CCCAGGCCCT CGCTGACGCC
4441 AGCCTGAGGA ACATGGTGCA GGCGGAGCAG GAGAACGACA CCTCGGGATG GTTTGATGTC
4501 CTCCAGAAAG TGTCTACCCA GTTGAAGACA AACCTCACGA GTGTCACAAA GAACCGTGCA
4561 GATAAGAATG CTATTCATAA TCACATTCGT TTGTTTGAAC CTCTTGTTAT AAAAGCTTTA
4621 AAACAGTACA CGACTACAAC ATGTGTGCAG TTACAGAAGC AGGTTTTAGA TTTGCTGGCG
4681 CAGCTGGTTC AGTTACGGGT TAATTACTGT CTTCTGGATT CAGATCAGGT GTTTATTGGC
4741 TTTGTATTGA AACAGTTTGA ATACATTGAA GTGGGCCAGT TCAGGGAATC AGAGGCAATC
4801 ATTCCAAACA TCTTTTTCTT CTTGGTATTA CTATCTTATG AACGCTATCA TTCAAAACAG
4861 ATCATTGGAA TTCCTAAAAT CATTCAGCTC TGTGATGGCA TCATGGCCAG TGGAAGGAAG
4921 GCTGTGACAC ATGCCATACC GGCTCTGCAG CCCATAGTCC ACGACCTCTT TGTATTAAGA
4981 GGAACAAATA AAGCTGATGC AGGAAAAGAG CTTGAAACCC AAAAAGAGGT GGTGGTGTCA
```

FIG.1D

```
5041 ATGTTACTGA GACTCATCCA GTACCATCAG GTGTTGGAGA TGTTCATTCT TGTCCTGCAG
5101 CAGTGCCACA AGGAGAATGA AGACAAGTGG AAGCGACTGT CTCGACAGAT AGCTGACATC
5161 ATCCTCCCAA TGTTAGCCAA ACAGCAGATG CACATTGACT CTCATGAAGC CCTTGGAGTG
5221 TTAAATACAT TATTTGAGAT TTTGGCCCCT TCCTCCCTCC GTCCGGTAGA CATGCTTTTA
5281 CGGAGTATGT TCGTCACTCC AAACACAATG GCGTCCGTGA GCACTGTTCA ACTGTGGATA
5341 TCGGGAATTC TGGCCATTTT GAGGGTTCTG ATTTCCCAGT CAACTGAAGA TATTGTTCTT
5401 TCTCGTATTC AGGAGCTCTC CTTCTCTCCG TATTTAATCT CCTGTACAGT AATTAATAGG
5461 TTAAGAGATG GGGACAGTAC TTCAACGCTA GAAGAACACA GTGAAGGGAA ACAAATAAAG
5521 AATTTGCCAG AAGAAACATT TTCAAGGTTT CTATTACAAC TGGTTGGTAT TCTTTTAGAA
5581 GACATTGTTA CAAAACAGCT GAAGGTGGAA ATGAGTGAGC AGCAACATAC TTTCTATTGC
5641 CAGGAACTAG GCACACTGCT AATGTGTCTG ATCCACATCT TCAAGTCTGG AATGTTCCGG
5701 AGAATCACAG CAGCTGCCAC TAGGCTGTTC CGCAGTGATG GCTGTGGCGG CAGTTTCTAC
5761 ACCCTGGACA GCTTGAACTT GCGGGCTCGT TCCATGATCA CCACCCACCC GGCCCTGGTG
5821 CTGCTCTGGT GTCAGATACT GCTGCTTGTC AACCACACCG ACTACCGCTG GTGGGCAGAA
5881 GTGCAGCAGA CCCCGAAAAG ACACAGTCTG TCCAGCACAA AGTTACTTAG TCCCCAGATG
5941 TCTGGAGAAG AGGAGGATTC TGACTTGGCA GCCAAACTTG GAATGTGCAA TAGAGAAATA
6001 GTACGAAGAG GGGCTCTCAT TCTCTTCTGT GATTATGTCT GTCAGAACCT CCATGACTCC
6061 GAGCACTTAA CGTGGCTCAT TGTAAATCAC ATTCAAGATC TGATCAGCCT TTCCCACGAG
6121 CCTCCAGTAC AGGACTTCAT CAGTGCCGTT CATCGGAACT CTGCTGCCAG CGGCCTGTTC
6181 ATCCAGGCAA TTCAGTCTCG TTGTGAAAAC CTTTCAACTC CAACCATGCT GAAGAAAACT
6241 CTTCAGTGCT TGGAGGGGAT CCATCTCAGC CAGTCGGGAG CTGTGCTCAC GCTGTATGTG
```

FIG.1E

```
6301 GACAGGCTTC TGTGCACCCC TTTCCGTGTG CTGGCTCGCA TGGTCGACAT CCTTGCTTGT
6361 CGCCGGGTAG AAATGCTTCT GGCTGCAAAT TTACAGAGCA GCATGGCCCA GTTGCCAATG
6421 GAAGAACTCA ACAGAATCCA GGAATACCTT CAGAGCAGCG GGCTCGCTCA GAGACACCAA
6481 AGGCTCTATT CCCTGCTGGA CAGGTTTCGT CTCTCCACCA TGCAAGACTC ACTTAGTCCC
6541 TCTCCTCCAG TCTCTTCCCA CCCGCTGGAC GGGGATGGGC ACGTGTCACT GGAAACAGTG
6601 AGTCCGGACA AAGACTGGTA CGTTCATCTT GTCAAATCCC AGTGTTGGAC CAGGTCAGAT
6661 TCTGCACTGC TGGAAGGTGC AGAGCTGGTG AATCGGATTC CTGCTGAAGA TATGAATGCC
6721 TTCATGATGA ACTCGGAGTT CAACCTAAGC CTGCTAGCTC CATGCTTAAG CCTAGGGATG
6781 AGTGAAATTT CTGGTGGCCA GAAGAGTGCC CTTTTTGAAG CAGCCCGTGA GGTGACTCTG
6841 GCCCGTGTGA GCGGCACCGT GCAGCAGCTC CCTGCTGTCC ATCATGTCTT CCAGCCCGAG
6901 CTGCCTGCAG AGCCGGCGGC CTACTGGAGC AAGTTGAATG ATCTGTTTGG GGATGCTGCA
6961 CTGTATCAGT CCCTGCCCAC TCTGGCCCGG GCCCTGGCAC AGTACCTGGT GGTGGTCTCC
7021 AAACTGCCCA GTCATTTGCA CCTTCCTCCT GAGAAAGAGA AGGACATTGT GAAATTCGTG
7081 GTGGCAACCC TTGAGGCCCT GTCCTGGCAT TTGATCCATG AGCAGATCCC GCTGAGTCTG
7141 GATCTCCAGG CAGGGCTGGA CTGCTGCTGC CTGGCCCTGC AGCTGCCTGG CCTCTGGAGC
7201 GTGGTCTCCT CCACAGAGTT TGTGACCCAC GCCTGCTCCC TCATCTACTG TGTGCACTTC
7261 ATCCTGGAGG CCGTTGCAGT GCAGCCTGGA GAGCAGCTTC TTAGTCCAGA AAGAAGGACA
7321 AATACCCCAA AAGCCATCAG CGAGGAGGAG GAGGAAGTAG ATCCAAACAC ACAGAATCCT
7381 AAGTATATCA CTGCAGCCTG TGAGATGGTG GCAGAAATGG TGGAGTCTCT GCAGTCGGTG
7441 TTGGCCTTGG GTCATAAAAG GAATAGCGGC GTGCCGGCGT TTCTCACGCC ATTGCTCAGG
7501 AACATCATCA TCAGCCTGGC CCGCCTGCCC CTTGTCAACA GCTACACACG TGTGCCCCCA
```

FIG.1F

```
7561 CTGGTGTGGA AGCTTGGATG GTCACCCAAA CCGGGAGGGG ATTTTGGCAC AGCATTCCCT
7621 GAGATCCCCG TGGAGTTCCT CCAGGAAAAG GAAGTCTTTA AGGAGTTCAT CTACCGCATC
7681 AACACACTAG GCTGGACCAG TCGTACTCAG TTTGAAGAAA CTTGGGCCAC CCTCCTTGGT
7741 GTCCTGGTGA CGCAGCCCCT CGTGATGGAG CAGGAGGAGA GCCCACCAGA AGAAGACACA
7801 GAGAGGACCC AGATCAACGT CCTGGCCGTG CAGGCCATCA CCTCACTGGT GCTCAGTGCA
7861 ATGACTGTGC CTGTGGCCGG CAACCCAGCT GTAAGCTGCT TGGAGCAGCA GCCCCGGAAC
7921 AAGCCTCTGA AAGCTCTCGA CACCAGGTTT GGGAGGAAGC TGAGCATTAT CAGAGGGATT
7981 GTGGAGCAAG AGATTCAAGC AATGGTTTCA AAGAGAGAGA ATATTGCCAC CCATCATTTA
8041 TATCAGGCAT GGGATCCTGT CCCTTCTCTG TCTCCGGCTA CTACAGGTGC CCTCATCAGC
8101 CACGAGAAGC TGCTGCTACA GATCAACCCC GAGCGGGAGC TGGGGAGCAT GAGCTACAAA
8161 CTCGGCCAGG TGTCCATACA CTCCGTGTGG CTGGGGAACA GCATCACACC CCTGAGGGAG
8221 GAGGAATGGG ACGAGGAAGA GGAGGAGGAG GCCGACGCCC CTGCACCTTC GTCACCACCC
8281 ACGTCTCCAG TCAACTCCAG GAAACACCGG GCTGGAGTTG ACATCCACTC CTGTTCGCAG
8341 TTTTTGCTTG AGTTGTACAG CCGCTGGATC CTGCCGTCCA GCTCAGCCAG GAGGACCCCG
8401 GCCATCCTGA TCAGTGAGGT GGTCAGATCC CTTCTAGTGG TCTCAGACTT GTTCACCGAG
8461 CGCAACCAGT TTGAGCTGAT GTATGTGACG CTGACAGAAC TGCGAAGGGT GCACCCTTCA
8521 GAAGACGAGA TCCTCGCTCA GTACCTGGTG CCTGCCACCT GCAAGGCAGC TGCCGTCCTT
8581 GGGATGGACA AGGCCGTGGC GGAGCCTGTC AGCCGCCTGC TGGAGAGCAC GCTCAGGAGC
8641 AGCCACCTGC CCAGCAGGGT TGGAGCCCTG CACGGCGTCC TCTATGTGCT GGAGTGCGAC
8701 CTGCTGGACG ACACTGCCAA GCAGCTCATC CCGGTCATCA GCGACTATCT CCTCTCCAAC
8761 CTGAAAGGGA TCGCCCACTG CGTGAACATT CACAGCCAGC AGCACGTACT GGTCATGTGT
```

FIG.1G

```
 8821 GCCACTGCGT TTTACCTCAT TGAGAACTAT CCTCTGGACG TAGGGCCGGA ATTTTCAGCA
 8881 TCAATAATAC AGATGTGTGG GGTGATGCTG TCTGGAAGTG AGGAGTCCAC CCCCTCCATC
 8941 ATTTACCACT GTGCCCTCAG AGGCCTGGAG CGCCTCCTGC TCTCTGAGCA GCTCTCCCGC
 9001 CTGGATGCAG AATCGCTGGT CAAGCTGAGT GTGGACAGAG TGAACGTGCA CAGCCCGCAC
 9061 CGGGCCATGG CGGCTCTGGG CCTGATGCTC ACCTGCATGT ACACAGGAAA GGAGAAAGTC
 9121 AGTCCGGGTA GAACTTCAGA CCCTAATCCT GCAGCCCCCG ACAGCGAGTC AGTGATTGTT
 9181 GCTATGGAGC GGGTATCTGT TCTTTTTGAT AGGATCAGGA AAGGCTTTCC TTGTGAAGCC
 9241 AGAGTGGTGG CCAGGATCCT GCCCCAGTTT CTAGACGACT TCTTCCCACC CCAGGACATC
 9301 ATGAACAAAG TCATCGGAGA GTTTCTGTCC AACCAGCAGC CATACCCCCA GTTCATGGCC
 9361 ACCGTGGTGT ATAAGGTGTT TCAGACTCTG CACAGCACCG GGCAGTCGTC CATGGTCCGG
 9421 GACTGGGTCA TGCTGTCCCT CTCCAACTTC ACGCAGAGGG CCCCGGTCGC CATGGCCACG
 9481 TGGAGCCTCT CCTGCTTCTT TGTCAGCGCG TCCACCAGCC CGTGGGTCGC GGCGATCCTC
 9541 CCACATGTCA TCAGCAGGAT GGGCAAGCTG GAGCAGGTGG ACGTGAACCT TTTCTGCCTG
 9601 GTCGCCACAG ACTTCTACAG ACACCAGATA GAGGAGGAGC TCGACCGCAG GGCCTTCCAG
 9661 TCTGTGCTTG AGGTGGTTGC AGCCCCAGGA AGCCCATATC ACCGGCTGCT GACTTGTTTA
 9721 CGAAATGTCC ACAAGGTCAC CACCTGC[redacted] GCGCCATGGT GGGAGAGACT GTGAGGCGGC
 9781 AGCTGGGGCC GGAGCCTTTG GAAGTCTGTG CCCTTGTGCC CTGCCTCCAC CGAGCCAGCT
 9841 TGGTCCCTAT GGGCTTCCGC ACATGCCGCG GGCGGCCAGG CAACGTGCGT GTCTCTGCCA
 9901 TGTGGCAGAA GTGCTCTTTG TGGCAGTGGC CAGGCAGGGA GTGTCTGCAG TCCTGGTGGG
 9961 GCTGAGCCTG AGGCCTTCCA GAAAGCAGGA GCAGCTGTGC TGCACCCCAT GTGGGTGACC
10021 AGGTCCTTTC TCCTGATAGT CACCTGCTGG TTGTTGCCAG GTTGCAGCTG CTCTTGCATC
```

FIG.1H

```
10081 TGGGCCAGAA GTCCTCCCTC CTGCAGGCTG GCTGTTGGCC CCTCTGCTGT CCTGCAGTAG
10141 AAGGTGCCGT GAGCAGGCTT TGGGAACACT GGCCTGGGTC TCCCTGGTGG GGTGTGCATG
10201 CCACGCCCCG TGTCTGGATG CACAGATGCC ATGGCCTGTG CTGGGCCAGT GGCTGGGGGT
10261 GCTAGACACC CGGCACCATT CTCCCTTCTC TCTTTTCTTC TCAGGATTTA AAATTTAATT
10321 ATATCAGTAA AGAGATTAAT TTTAACGAAC TCTTTCTATG CCCGTGTAAA GTATGTGAAT
10381 CGCAAGGCCT GTGCTGCATG CGACAGCGTC CGGGGTGGTG GACAGGGCCC CCGGCCACGC
10441 TCCCTCTCCT GTAGCCACTG GCATAGCCCT CCTGAGCACC CGCTGACATT TCCGTTGTAC
10501 ATGTTCCTGT TTATGCATTC ACAAGGTGAC TGGGATGTAG AGAGGCGTTA GTGGGCAGGT
10561 GGCCACAGCA GGACTGAGGA CAGGCCCCCA TTATCCTAGG GGTGCGCTCA ACTGCAGCCC
10621 CTCCTCCTCG GGCACAGACG ACTGTCGTTC TCCACCCACC AGTCAGGGAC AGCAGCCTCC
10681 CTGTCACTCA GCTGAGAAGG CCAGCCCTCC CTGGCTGTGA GCAGCCTCCA CTGTGTCCAG
10741 AGACATGGGC CTCCCACTCC TGTTCCTTGC TAGCCCTGGG GTGGCGTCTG CCTAGGAGCT
10801 GGCTGGCAGG TGTTGGGACC TGCTGCTCCA TGGATGCATG CCCTAAGAGT GTCACTGAGC
10861 TGTGTTTTGT CTGAGCCTCT CTCGGTCAAC AGCAAAGCTT GGTGTCTTGG CACTGTTAGT
10921 GACAGAGCCC AGCATCCCTT CTGCCCCCGT TCCAGCTGAC ATCTTGCACG GTGACCCCTT
10981 TTAGTCAGGA GAGTGCAGAT CTGTGCTCAT CGGAGACTGC CCCACGGCCC TGTCAGAGCC
11041 GCCACTCCTA TCCCCAGGAC AGGTCCCTGG ACCAGCCTCC TGTTTGCAGG CCCAGAGGAG
11101 CCAAGTCATT AAAATGGAAG TGGATTCTGG ATGGCCGGGC TGCTGCTGAT GTAGGAGCTG
11161 GATTTGGGAG CTCTGCTTGC CGACTGGCTG TGAGACGAGG CAGGGGCTCT GCTTCCTCAG
11221 CCCTAGAGGC GAGCCAGGCA AGGTTGGCGA CTGTCATGTG GCTTGGTTTG GTCATGCCCG
11281 TCGATGTTTT GGGTATTGAA TGTGGTAAGT GGAGGAAATG TTGGAACTCT GTGCAGGTGC
```

FIG.1I

```
11341 TGCCTTGAGA CCCCCAAGCT TCCACCTGTC CCTCTCCTAT GTGGCAGCTG GGGAGCAGCT
11401 GAGATGTGGA CTTGTATGCT GCCCACATAC GTGAGGGGGA GCTGAAAGGG AGCCCCTGCT
11461 CAAAGGGAGC CCCTCCTCTG AGCAGCCTCT GCCAGGCCTG TATGAGGCTT TTCCCACCAG
11521 CTCCCAACAG AGGCCTCCCC CAGCCAGGAC CACCTCGTCC TCGTGGCGGG GCAGCAGGAG
11581 CGGTAGAAAG GGGTCCGATG TTTGAGGAGG CCCTTAAGGG AAGCTACTGA ATTATAACAC
11641 GTAAGAAAAT CACCATTCTT CCGTATTGGT TGGGGGCTCC TGTTTCTCAT CCTAGCTTTT
11701 TCCTGGAAAA GCCCGCTAGA AGGTTTGGGA ACGAGGGGAA AGTTCTCAGA ACTGTTGCTG
11761 CTCCCCACCC GCCTCCCGCC TCCCCCGCAG GTTATGTCAG CAGCTCTGAG ACAGCAGTAT
11821 CACAGGCCAG ATGTTGTTCC TGGCTAGATG TTTACATTTG TAAGAAATAA CACTGTGAAT
11881 GTAAAACAGA GCCATTCCCT TGGAATGCAT ATCGCTGGGC TCAACATAGA GTTTGTCTTC
11941 CTCTTGTTTA CGACGTGATC TAAACCAGTC CTTAGCAAGG GGCTCAGAAC ACCCCGCTCT
12001 GGCAGTAGGT GTCCCCCACC CCCAAAGACC TGCCTGTGTG CTCCGGAGAT GAATATGAGC
12061 TCATTAGTAA AAATGACTTC ACCCACGCAT ATACATAAAG TATCCATGCA TGTGCATATA
12121 GACACATCTA TAATTTTACA CACACACCTC TCAAGACGGA GATGCATGGC CTCTAAGAGT
12181 GCCCGTGTCG GTTCTTCCTG GAAGTTGACT TTCCTTAGAC CCGCCAGGTC AAGTTAGCCG
12241 CGTGACGGAC ATCCAGGCGT GGGACGTGGT CAGGGCAGGG CTCATTCATT GCCCACTAGG
12301 ATCCCACTGG CGAAGATGGT CTCCATATCA GCTCTCTGCA GAAGGGAGGA AGACTTTATC
12361 ATGTTCCTAA AAATCTGTGG CAAGCACCCA TCGTATTATC CAAATTTTGT TGCAAATGTG
12421 ATTAATTTGG TTGTCAAGTT TTGGGGGTGG GCTGTGGGGA GATTGCTTTT GTTTTCCTGC
12481 TGGTAATATC GGGAAAGATT TTAATGAAAC CAGGGTAGAA TTGTTTGGCA ATGCACTGAA
12541 GCGTGTTTCT TTCCCAAAAT GTGCCTCCCT TCCGCTGCGG GCCCAGCTGA GTCTATGTAG
```

FIG.1J

```
12601 GTGATGTTTC CAGCTGCCAA GTGCTCTTTG TTACTGTCCA CCCTCATTTC TGCCAGCGCA
12661 TGTGTCCTTT CAAGGGGAAA ATGTGAAGCT GAACCCCCTC CAGACACCCA GAATGTAGCA
12721 TCTGAGAAGG CCCTGTGCCC TAAAGGACAC CCCTCGCCCC CATCTTCATG GAGGGGGTCA
12781 TTTCAGAGCC CTCGGAGCCA ATGAACAGCT CCTCCTCTTG GAGCTGAGAT GAGCCCCACG
12841 TGGAGCTCGG GACGGATAGT AGACAGCAAT AACTCGGTGT GTGGCCGCCT GGCAGGTGGA
12901 ACTTCCTCCC GTTGCGGGGT GGAGTGAGGT TAGTTCTGTG TGTCTGGTGG GTGGAGTCAG
12961 GCTTCTCTTG CTACCTGTGA GCATCCTTCC CAGCAGACAT CCTCATCGGG CTTTGTCCCT
13021 CCCCCGCTTC CTCCCTCTGC GGGGAGGACC CGGGACCACA GCTGCTGGCC AGGGTAGACT
13081 TGGAGCTGTC CTCCAGAGGG GTCACGTGTA GGAGTGAGAA GAAGGAAGAT CTTGAGAGCT
13141 GCTGAGGGAC CTTGGAGAGC TCAGGATGGC TCAGACGAGG ACACTCGCTT GCCGGGCCTG
13201 GCCCTCCTGG GAAGGAGGGA GCTGCTCAGA ATGCCGCATG ACAACTGAAG GCAACCTGGA
13261 AGGTTCAGGG CCCGCTCTTC CCCCATGTGC CTGTCACGCT CTGGTGCAGT CAAAGGAACG
13321 CCTTCCCCTC AGTTGTTTCT AAGAGCAGAG TCTCCCGCTG CAATCTGGGT GGTAACTGCC
13381 AGCCTTGGAG GATCGTGGCC AACGTGGACC TGCCTACGGA GGGTGGGCTC TGACCCAAGT
13441 GGGGCCTCCT TGCCCAGGTC TCACTGCTTT GCACCGTGGT CAGAGGGACT GTCAGCTGAG
13501 CTTGAGCTCC CCTGGAGCCA GCAGGGCTGT GATGGGCGAG TCCCGGAGCC CCACCCAGAC
13561 CTGAATGCTT CTGAGAGCAA AGGGAAGGAC TGACGAGAGA TGTATATTTA ATTTTTTAAC
13621 TGCTGCAAAC ATTGTACATC CAAATTAAAG GGAAAAAATG GAAACCATCA AT
```

FIG.1K

FIG.2A

```
   1 matleklmka feslksfqqq qqqqqqqqqq qqqqqqqqqq ppppppppqp pqlpqpppqa
  61 qpllpqpqpp pppppppqgp avaeeplhrp kkelsatkkd rvnhcltice nivaqsvrns
 121 pefqkllgia melfllcsdd aesdvrmvad eclnkvikal mdsnlprlql elykeikkng
 181 aprslraalw rfaelahlvr pqkcrpylvn llpcltrtsk rpeesvqetl aaavpkimas
 241 fgnfandnei kvllkafian lkssgptirr taagsavsic qhsrrtqyfy swllnvllgl
 301 lvpvedehst llilgvlltl rylvpllqqq vkdtslkgsf gvtrkemevs psaeqlvqvy
 361 eltlhhtqhq dhnvvtgale llqqlfrtpp pellqtltav ggigqltaak eesggrsrsg
 421 siveliaggg sscspvlsrk qkgkvllgee ealeddsesr sdvsssalta svkdeisgel
 481 aassgvstpg saghditeq prsqhtlqad svdlascdlt ssatdgdeed ilshsssqvs
 541 avpsdpamdl ndgtqasspi sdssqtttеg pdsavtpsds seivldgtdn qylglqigqp
 601 qdedeeatgi lpdeaseafr nssmalqqah llknmshcrq psdssvdkfv lrdeatepgd
 661 qenkpcrikg diggstddds aplvhcvrll sasfiltggk nvlvpdrdvr vsvkalalsc
 721 vgaavalhpe sffsklykvp ldtteypeeq yvsdilnyid hgdpqvrgat ailcgtlics
 781 ilsrsrfhvg dwmgtirtlt gntfsladci pllrktlkde ssvtcklact avrncvmslc
 841 sssyselglq liidvltlrn ssywlvrtel letlaeidfr lvsfleakae nlhrgahhyt
 901 gllklqervl nnvvihllgd edprvrhvaa aslirlvpkl fykcdqgqad pvvavardqs
 961 svylkllmhe tqppshfsvs titriyrgyn llpsitdvtm ennlsrviaa vshelitstt
1021 raltfgccea lcllstafpv ciwslgwhcg vpplsasdes rksctvgmat milтllssaw
1081 fpldlsahqd alilagnlla asapkslrss waseeeanpa atkqeevwpa lgdralvpmv
1141 eqlfshllkv inicahvldd vapgpaikaa lpsltnppsl spirrkgkek epgeqasvpl
1201 spkkgseasa asrqsdtsgp vttskssslg sfyhlpsylk lhdvlkatha nykvtldlqn
1261 stekfggflr saldvlsqil elatlqdigk cveeilgylk scfsrepmma tvcvqqllkt
1321 lfgtnlasqf dglssnpsks qgraqrlgss svrpglyhyc fmapythftq aladaslrnm
1381 vqaeqendts gwfdvlqkvs tqlktnltsv tknradknai hnhirlfepl vikalkqytt
1441 ttcvqlqkqv ldllaqlvql rvnyclldsd qvfigfvlkq feyievgqfr eseaiipnif
1501 fflvllsyer yhskqiigip kiiqlcdgim asgrkavtha ipalqpivhd lfvlrgtnka
1561 dagkeletqk evvvsmllrl iqyhqvlemf ilvlqqchke nedkwkrlsr qiadiilpml
1621 akqqmhidsh ealgvlntlf eilapsslrp vdmllrsmfv tpntmasvst vqlwisgila
1681 ilrvlisqst edivlsriqe lsfspylisc tvinrlrdgd ststleehse gkqiknlpee
```

FIG.2B

```
1741 tfsrfllqlv gilledivtk qlkvemseqq htfycqelgt llmclihifk sgmfrritaa
1801 atrlfrsdgc ggsfytldsl nlrarsmitt hpalvllwcq illlvnhtdy rwwaevqgtp
1861 krhslsstkl lspqmsgeee dsdlaaklgm cnreivrrga lilfcdyvcq nlhdsehltw
1921 livnhiqdli slsheppvqd fisavhrnsa asglfiqaiq srcenlstpt mlkktlqcle
1981 gihlsqsgav ltlyvdrllc tpfrvlarmv dilacrrvem llaanlqssm aqlpmeelnr
2041 iqeylqssgl aqrhqrlysl ldrfrlstmq dslspsppvs shpldgdghv sletvspdkd
2101 wyvhlvksqc wtrsdsalle gaelvnripa edmnafmmns efnlsllapc lslgmseisg
2161 gqksalfeaa revtlarvsg tvqqlpavhh vfqpelpaep aaywsklndl fgdaalyqsl
2221 ptlaralaqy lvvvsklpsh lhlppekekd ivkfvvatle alswhliheq iplsldlqag
2281 ldccclalql pglwsvvsst efvthacsli ycvhfileav avqpgeqlls perrtntpka
2341 iseeeeevdp ntqmpkyita acemvaemve slqsvlalgh krnsgvpafl tpllrniiis
2401 larlplvnsy trvpplvwkl gwspkpggdf gtafpeipve flqekevfke fiyrintlgw
2461 tsrtqfeetw atllgvlvtq plvmeqeesp peedtertqi nvlavqaits lvlsamtvpv
2521 agnpavscle qqprnkplka ldtrfgrkls iirgiveqei qamvskreni athhlyqawd
2581 pvpslspatt galishekll lqinperelg smsyklgqvs ihsvwlgnsi tplreeewde
2641 eeeeeadapa pspptspvn srkhragvdi hscsqfllel ysrwilpsss arrtpailis
2701 evvrsllvvs dlfternqfe lmyvtltelr rvhpsedeil aqylvpatck aaavlgmdka
2761 vaepvsrlle stlrsshlps rvgalhgvly vlecdllddt akqlipvisd yllsnlkgia
2821 hcvnihsqqh vlvmcatafy lienypldvg pefsasiiqm cgvmlsgsee stpsiiyhca
2881 lrglerllls eqlsrldaes lvklsvdrvn vhsphramaa lglmltcmyt gkekvspgrt
2941 sdpnpaapds esvivamerv svlfdrirkg fpcearvvar ilpqflddff ppqdimnkvi
3001 geflsnqqpy pqfmatvvyk vfqtlhstgq ssmvrdwvml slsnftqrap vamatwslsc
3061 ffvsastspw vaailphvis rmgkleqvdv nlfclvatdf yrhqieeeld rrafqsvlev
3121 vaapgspyhr lltclrnvhk vttc
```

FIG. 3

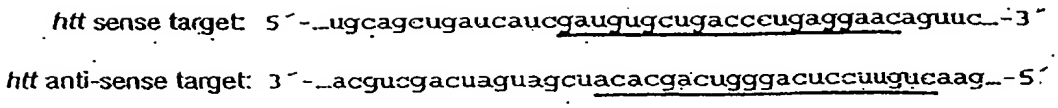

htt sense target: 5'-ugcagcugaucaucgaugugcugacccugaggaacaguuc-3'
htt anti-sense target: 3'-acgucgacuaguagcuacacgacugggacuccuugucaag-5'

FIG. 4

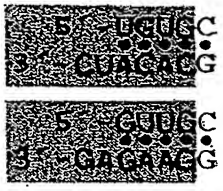


FIG. 5A

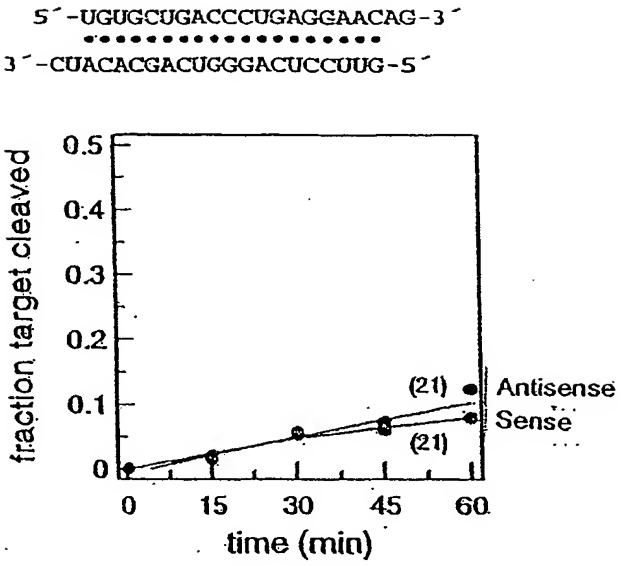

5´-UGUGCUGACCCUGAGGAACAG-3´
3´-CUACACGACUGGGACUCCUUG-5´
fraction target cleaved
0.5
0.4
0.3
0.2
0.1
0
0
15
30
45
60
time (min)
(21)
Antisense
Sense
(21)

FIG. 5B

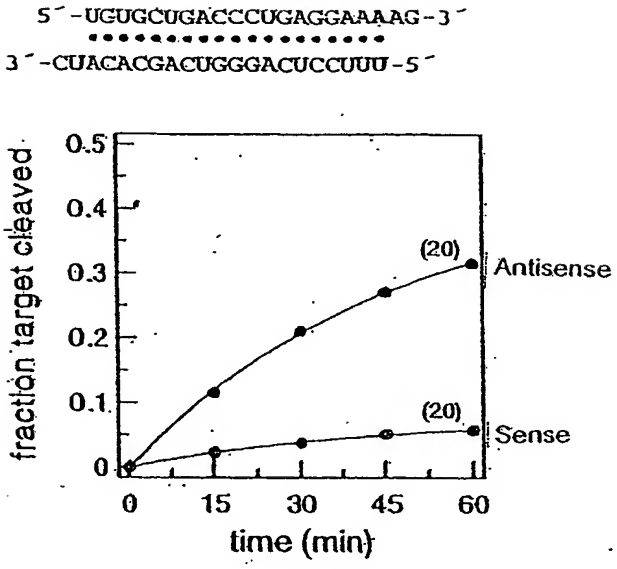

5´-UGUGCUGACCCUGAGGAAAAG-3´
3´-CUACACGACUGGGACUCCUUU-5´
fraction target cleaved
0.5
0.4
0.3
0.2
0.1
0
(20)
Antisense
(20)
Sense
0
15
30
45
60
time (min)

FIG. 5C

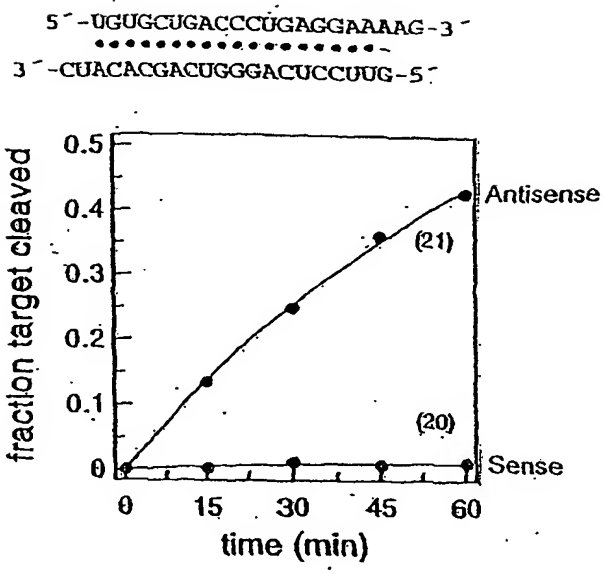
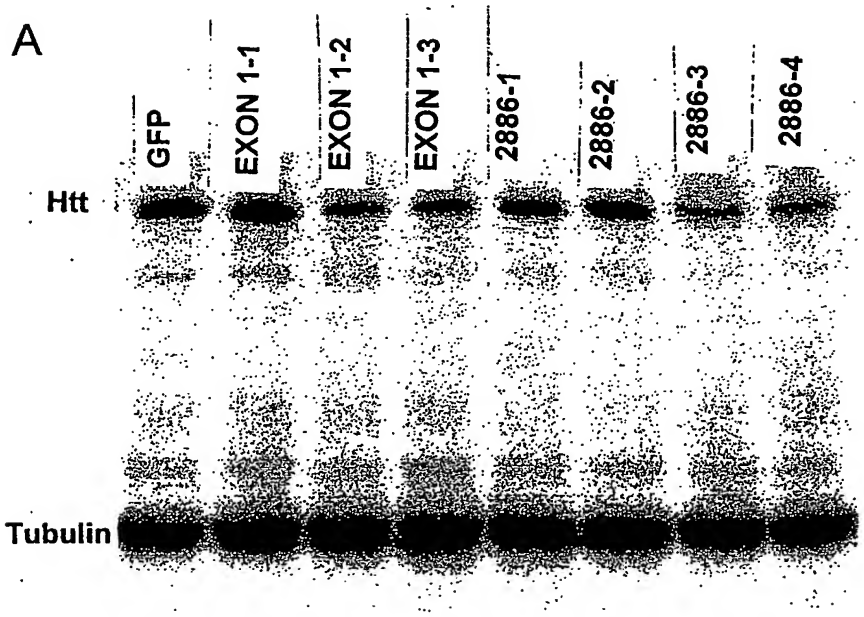
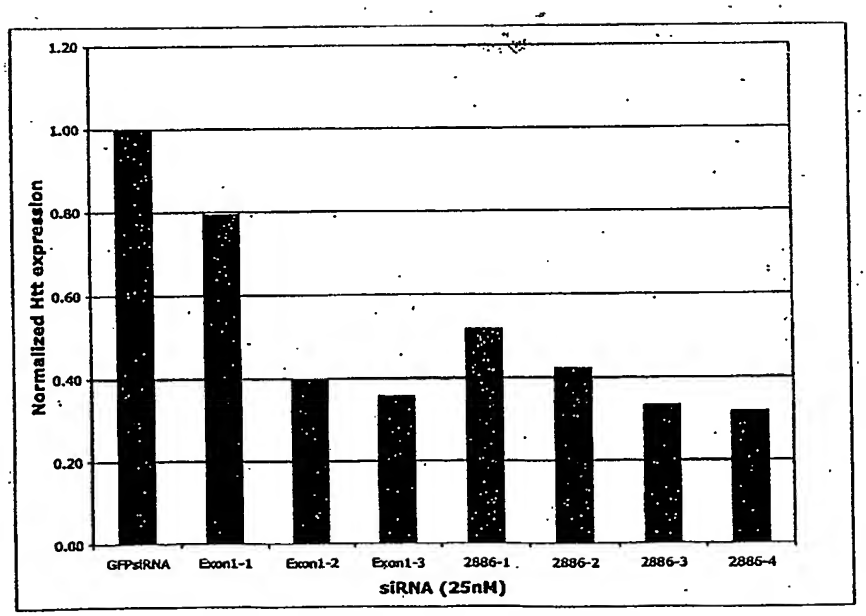

5´-UGUGCUGACCCUGAGGAAAAG-3´
3´-CUACACGACUGGGACUCCUUG-5´
fraction target cleaved
0.5
0.4
0.3
0.2
0.1
0
0
15
30
45
60
time (min)
Antisense
(21)
(20)
Sense

FIG. 6


A
GFP
EXON 1-1
EXON 1-2
EXON 1-3
2886-1
2886-2
2886-3
2886-4
Htt
Tubulin

B


1.20
1.00
0.80
0.60
0.40
0.20
0.00
Normalized Htt expression
GFPsiRNA
Exon1-1
Exon1-2
Exon1-3
2886-1
2886-2
2886-3
2886-4
siRNA (25nM)

SEQUENCE LISTING

<110> University of Massachusetts
ARONIN, Neil
ZAMORE, Phillip D.

<120> RNA INTERFERENCE FOR THE TREATMENT OF GAIN-OF-FUNCTION DISORDERS

<130> UMY-083PC

<150> 60/502678
<151> 2003-09-12

<160> 33

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 13672
<212> DNA
<213> Homo sapiens

<400> 1
ttgctgtgtg aggcagaacc tgcgggggca ggggcgggct ggttccctgg ccagccattg 60
gcagagtccg caggctaggg ctgtcaatca tgctggccgg cgtggccccg cctccgccgg 120
cgcggccccg cctccgccgg cgcacgtctg ggacgcaagg cgccgtgggg gctgccggga 180
cgggtccaag atggacggcc gctcaggttc tgcttttacc tgcggcccag agccccattc 240
attgccccgg tgctgagcgg cgccgcgagt cggcccgagg cctccgggga ctgccgtgcc 300
gggcgggaga ccgccatggc gaccctggaa aagctgatga aggccttcga gtccctcaag 360
tccttccagc agcagcagca gcagcagcag cagcagcagc agcagcagca gcagcagcag 420
cagcagcagc aacagccgcc accgccgccg ccgccgccgc cgcctcctca gcttcctcag 480
ccgccgccgc aggcacagcc gctgctgcct cagccgcagc cgcccccgcc gccgcccccg 540
ccgccacccg gcccggctgt ggctgaggag ccgctgcacc gaccaaagaa agaactttca 600
gctaccaaga aagaccgtgt gaatcattgt ctgacaatat gtgaaaacat agtggcacag 660
tctgtcagaa attctccaga atttcagaaa cttctgggca tcgctatgga acttttctg 720
ctgtgcagtg atgacgcaga gtcagatgtc aggatggtgg ctgacgaatg cctcaacaaa 780
gttatcaaag ctttgatgga ttctaatctt ccaaggttac agctcgagct ctataaggaa 840
attaaaaaga atggtgcccc tcggagtttg cgtgctgccc tgtggaggtt tgctgagctg 900
gctcacctgg ttcggcctca gaaatgcagg ccttacctgg tgaaccttct gccgtgcctg 960
actcgaacaa gcaagagacc cgaagaatca gtccaggaga ccttggctgc agctgttccc 1020
aaaattatgg cttcttttgg caattttgca aatgacaatg aaattaaggt tttgttaaag 1080
gccttcatag cgaacctgaa gtcaagctcc cccaccattc ggcggacagc ggctggatca 1140
gcagtgagca tctgccagca ctcaagaagg acacaatatt tctatagttg gctactaaat 1200
gtgctcttag gcttactcgt tcctgtcgag gatgaacact ccactctgct gattcttggc 1260
gtgctgctca ccctgaggta tttggtgccc ttgctgcagc agcaggtcaa ggacacaagc 1320
ctgaaaggca gcttcggagt gacaaggaaa gaaatggaag tctctccttc tgcagagcag 1380
cttgtccagg tttatgaact gacgttacat catacacagc accaagacca caatgttgtg 1440
accggagccc tggagctgtt gcagcagctc ttcagaacgc ctccacccga gcttctgcaa 1500
accctgaccg cagtcggggg cattgggcag ctcaccgctg ctaaggagga gtctggtggc 1560
cgaagccgta gtgggagtat tgtggaactt atagctggag ggggttcctc atgcagccct 1620
gtcctttcaa gaaaacaaaa aggcaaagtg ctcttaggag aagaagaagc cttggaggat 1680
gactctgaat cgagatcgga tgtcagcagc tctgccttaa cagcctcagt gaaggatgag 1740
atcagtggag agctggctgc ttcttcaggg gtttccactc cagggtcagc aggtcatgac 1800
atcatcacag aacagccacg gtcacagcac acactgcagg cggactcagt ggatctggcc 1860
agctgtgact tgacaagctc tgccactgat ggggatgagg aggatatctt gagccacagc 1920
tccagccagg tcagcgccgt cccatctgac cctgccatgg acctgaatga tgggacccag 1980
gcctcgtcgc ccatcagcga cagctcccag accaccaccg aagggcctga ttcagctgtt 2040
acccccttcag acagttctga aattgtgtta gacggtaccg acaaccagta tttgggcctg 2100

cagattggac agccccagga tgaagatgag gaagccacag gtattcttcc tgatgaagcc 2160
tcggaggcct tcaggaactc ttccatggcc cttcaacagg cacatttatt gaaaaacatg 2220
agtcactgca ggcagccttc tgacagcagt gttgataaat ttgtgttgag agatgaagct 2280
actgaaccgg gtgatcaaga aaacaagcct tgccgcatca aaggtgacat tggacagtcc 2340
actgatgatg actctgcacc tcttgtccat tgtgtccgcc ttttatctgc ttcgtttttg 2400
ctaacagggg gaaaaaatgt gctggttccg gacagggatg tgagggtcag cgtgaaggcc 2460
ctggccctca gctgtgtggg agcagctgtg gccctccacc cggaatcttt cttcagcaaa 2520
ctctataaag ttcctcttga caccacggaa taccctgagg aacagtatgt ctcagacatc 2580
ttgaactaca tcgatcatgg agacccacag gttcgaggag ccactgccat tctctgtggg 2640
accctcatct gctccatcct cagcaggtcc cgcttccacg tgggagattg gatgggcacc 2700
attagaaccc tcacaggaaa tacattttct ttggcggatt gcattccttt gctgcggaaa 2760
acactgaagg atgagtcttc tgttacttgc aagttagctt gtacagctgt gaggaactgt 2820
gtcatgagtc tctgcagcag cagctacagt gagttaggac tgcagctgat catcgatgtg 2880
ctgactctga ggaacagttc ctattggctg gtgaggacag agcttctgga aacccttgca 2940
gagattgact tcaggctggt gagcttttg gaggcaaaag cagaaaactt acacagaggg 3000
gctcatcatt atacagggct tttaaaactg caagaacgag tgctcaataa tgttgtcatc 3060
catttgcttg gagatgaaga ccccagggtg cgacatgttg ccgcagcatc actaattagg 3120
cttgtcccaa agctgtttta taaatgtgac caaggacaag ctgatccagt agtggccgtg 3180
gcaagagatc aaagcagtgt ttacctgaaa cttctcatgc atgagacgca gcctccatct 3240
catttctccg tcagcacaat aaccagaata tatagaggct ataacctact accaagcata 3300
acagacgtca ctatggaaaa taacctttca agagttattg cagcagtttc tcatgaacta 3360
atcacatcaa ccaccagagc actcacattt ggatgctgtg aagctttgtg tcttctttcc 3420
actgccttcc cagtttgcat ttggagttta ggttggcact gtggagtgcc tccactgagt 3480
gcctcagatg agtctaggaa gagctgtacc gttgggatgg ccacaatgat tctgaccctg 3540
ctctcgtcag cttggttccc attggatctc tcagcccatc aagatgcttt gattttggcc 3600
ggaaacttgc ttgcagccag tgctcccaaa tctctgagaa gttcatgggc ctctgaagaa 3660
gaagccaacc cagcagccac caagcaagag gaggtctggc cagccctggg ggaccgggcc 3720
ctggtgccca tggtggagca gctcttctct cacctgctga aggtgattaa catttgtgcc 3780
cacgtcctgg atgacgtggc tcctggaccc gcaataaagg cagccttgcc ttctctaaca 3840
aacccccctt ctctaagtcc catccgacga aaggggaagg agaaagaacc aggagaacaa 3900
gcatctgtac cgttgagtcc caagaaaggc agtgaggcca gtgcagcttc tagacaatct 3960
gatacctcag gtcctgttac aacaagtaaa tcctcatcac tggggagttt ctatcatctt 4020
ccttcatacc tcaaactgca tgatgtcctg aaagctacac acgctaacta caaggtcacg 4080
ctggatcttc agaacagcac ggaaaagttt ggagggtttc tccgctcagc cttggatgtt 4140
ctttctcaga tactagagct ggccacactg caggacattg ggaagtgtgt tgaagagatc 4200
ctaggatacc tgaaatcctg ctttagtcga gaaccaatga tggcaactgt ttgtgttcaa 4260
caattgttga agactctctt tggcacaaac ttggcctccc agtttgatgg cttatcttcc 4320
aaccccagca agtcacaagg ccgagcacag cgccttggct cctccagtgt gaggccaggc 4380
ttgtaccact actgcttcat ggccccgtac acccacttca cccaggccct cgctgacgcc 4440
agcctgagga acatggtgca ggcggagcag gagaacgaca cctcgggatg gtttgatgtc 4500
ctccagaaag tgtctaccca gttgaagaca aacctcacga gtgtcacaaa gaaccgtgca 4560
gataagaatg ctattcataa tcacattcgt ttgtttgaac ctcttgttat aaaagcttta 4620
aaacagtaca cgactacaac atgtgtgcag ttacagaagc aggttttaga tttgctggcg 4680
cagctggttc agttacgggt taattactgt cttctggatt cagatcaggt gtttattggc 4740
tttgtattga aacagtttga atacattgaa gtgggccagt tcagggaatc agaggcaatc 4800
attccaaaca tctttttctt cttggtatta ctatcttatg aacgctatca ttcaaaacag 4860
atcattggaa ttcctaaaat cattcagctc tgtgatggca tcatggccag tggaaggaag 4920
gctgtgacac atgccatacc ggctctgcag cccatagtcc acgacctctt tgtattaaga 4980
ggaacaaata aagctgatgc aggaaaagag cttgaaaccc aaaaagaggt ggtggtgtca 5040
atgttactga gactcatcca gtaccatcag gtgttggaga tgttcattct tgtcctgcag 5100
cagtgccaca aggagaatga agacaagtgg aagcgactgt ctcgacagat agctgacatc 5160
atcctcccaa tgttagccaa acagcagatg cacattgact ctcatgaagc ccttggagtg 5220
ttaaatacat tatttgagat tttggcccct tcctccctcc gtccggtaga catgctttta 5280
cggagtatgt tcgtcactcc aaacacaatg gcgtccgtga gcactgttca actgtggata 5340
tcgggaattc tggccatttt gagggttctg atttcccagt caactgaaga tattgttctt 5400
tctcgtattc aggagctctc cttctctccg tatttaatct cctgtacagt aattaatagg 5460
ttaagagatg gggacagtac ttcaacgcta gaagaacaca gtgaagggaa acaaataaag 5520
aatttgccag aagaaacatt ttcaaggttt ctattacaac tggttggtat tcttttagaa 5580
gacattgtta caaaacagct gaaggtggaa atgagtgagc agcaacatac tttctattgc 5640
caggaactag gcacactgct aatgtgtctg atccacatct tcaagtctgg aatgttccgg 5700
agaatcacag cagctgccac taggctgttc cgcagtgatg gctgtggcgg cagtttctac 5760

accctggaca gcttgaactt gcgggctcgt tccatgatca ccacccaccc ggccctggtg 5820
ctgctctggt gtcagatact gctgcttgtc aaccacaccg actaccgctg gtgggcagaa 5880
gtgcagcaga ccccgaaaag acacagtctg tccagcacaa agttacttag tccccagatg 5940
tctggagaag aggaggattc tgacttggca gccaaacttg gaatgtgcaa tagagaaata 6000
gtacgaagag gggctctcat tctcttctgt gattatgtct gtcagaacct ccatgactcc 6060
gagcacttaa cgtggctcat tgtaaatcac attcaagatc tgatcagcct ttcccacgag 6120
cctccagtac aggacttcat cagtgccgtt catcggaact ctgctgccag cggcctgttc 6180
atccaggcaa ttcagtctcg ttgtgaaaac ctttcaactc caaccatgct gaagaaaact 6240
cttcagtgct tggaggggat ccatctcagc cagtcgggag ctgtgctcac gctgtatgtg 6300
gacaggcttc tgtgcacccc tttccgtgtg ctggctcgca tggtcgacat ccttgcttgt 6360
cgccgggtag aaatgcttct ggctgcaaat ttacagagca gcatggccca gttgccaatg 6420
gaagaactca acagaatcca ggaatacctt cagagcagcg ggctcgctca gagacaccaa 6480
aggctctatt ccctgctgga caggtttcgt ctctccacca tgcaagactc acttagtccc 6540
tctcctccag tctcttccca cccgctggac ggggatgggc acgtgtcact ggaaacagtg 6600
agtccggaca aagactggta cgttcatctt gtcaaatccc agtgttggac caggtcagat 6660
tctgcactgc tggaaggtgc agagctggtg aatcggattc ctgctgaaga tatgaatgcc 6720
ttcatgatga actcggagtt caacctaagc ctgctagctc catgcttaag cctagggatg 6780
agtgaaattt ctggtggcca gaagagtgcc ctttttgaag cagcccgtga ggtgactctg 6840
gcccgtgtga gcggcaccgt gcagcagctc cctgctgtcc atcatgtctt ccagcccgag 6900
ctgcctgcag agccggcggc ctactggagc aagttgaatg atctgtttgg ggatgctgca 6960
ctgtatcagt ccctgcccac tctggcccgg gccctggcac agtacctggt ggtggtctcc 7020
aaactgccca gtcatttgca ccttcctcct gagaaagaga aggacattgt gaaattcgtg 7080
gtggcaaccc ttgaggccct gtcctggcat ttgatccatg agcagatccc gctgagtctg 7140
gatctccagg cagggctgga ctgctgctgc ctggccctgc agctgcctgg cctctggagc 7200
gtggtctcct ccacagagtt tgtgacccac gcctgctccc tcatctactg tgtgcacttc 7260
atcctggagg ccgttgcagt gcagcctgga gagcagcttc ttagtccaga aagaaggaca 7320
aataccccaa aagccatcag cgaggaggag gaggaagtag atccaaacac acagaatcct 7380
aagtatatca ctgcagcctg tgagatggtg gcagaaatgg tggagtctct gcagtcggtg 7440
ttggccttgg gtcataaaag gaatagcggc gtgccggcgt ttctcacgcc attgctcagg 7500
aacatcatca tcagcctggc ccgcctgccc cttgtcaaca gctacacacg tgtgccccca 7560
ctggtgtgga agcttggatg gtcacccaaa ccgggagggg attttggcac agcattccct 7620
gagatccccg tggagttcct ccaggaaaag gaagtcttta aggagttcat ctaccgcatc 7680
aacacactag gctggaccag tcgtactcag tttgaagaaa cttgggccac cctccttggt 7740
gtcctggtga cgcagcccct cgtgatggag caggaggaga gcccaccaga agaagacaca 7800
gagaggaccc agatcaacgt cctggccgtg caggccatca cctcactggt gctcagtgca 7860
atgactgtgc ctgtggccgg caacccagct gtaagctgct tggagcagca gccccggaac 7920
aagcctctga aagctctcga caccaggttt gggaggaagc tgagcattat cagagggatt 7980
gtggagcaag agattcaagc aatggtttca aagagagaga atattgccac ccatcattta 8040
tatcaggcat gggatcctgt cccttctctg tctccggcta ctacaggtgc cctcatcagc 8100
cacgagaagc tgctgctaca gatcaacccc gagcgggagc tggggagcat gagctacaaa 8160
ctcggccagg tgtccataca ctccgtgtgg ctggggaaca gcatcacacc cctgagggag 8220
gaggaatggg acgaggaaga ggaggaggag gccgacgccc ctgcaccttc gtcaccaccc 8280
acgtctccag tcaactccag gaaacaccgg gctggagttg acatccactc ctgttcgcag 8340
tttttgcttg agttgtacag ccgctggatc ctgccgtcca gctcagccag gaggaccccg 8400
gccatcctga tcagtgaggt ggtcagatcc cttctagtgg tctcagactt gttcaccgag 8460
cgcaaccagt ttgagctgat gtatgtgacg ctgacagaac tgcgaagggt gcacccttca 8520
gaagacgaga tcctcgctca gtacctggtg cctgccacct gcaaggcagc tgccgtcctt 8580
gggatggaca aggccgtggc ggagcctgtc agccgcctgc tggagagcac gctcaggagc 8640
agccacctgc ccagcagggt tggagccctg cacggcgtcc tctatgtgct ggagtgcgac 8700
ctgctggacg acactgccaa gcagctcatc ccggtcatca gcgactatct cctctccaac 8760
ctgaaaggga tcgcccactg cgtgaacatt cacagccagc agcacgtact ggtcatgtgt 8820
gccactgcgt tttacctcat tgagaactat cctctggacg tagggccgga attttcagca 8880
tcaataatac agatgtgtgg ggtgatgctg tctggaagtg aggagtccac cccctccatc 8940
atttaccact gtgccctcag aggcctggag cgcctcctgc tctctgagca gctctcccgc 9000
ctggatgcag aatcgctggt caagctgagt gtggacagag tgaacgtgca cagcccgcac 9060
cgggccatgg cggctctggg cctgatgctc acctgcatgt acacaggaaa ggagaaagtc 9120
agtccgggta gaacttcaga ccctaatcct gcagcccccg acagcgagtc agtgattgtt 9180
gctatggagc gggtatctgt tctttttgat aggatcagga aaggctttcc ttgtgaagcc 9240
agagtggtgg ccaggatcct gccccagttt ctagacgact tcttcccacc ccaggacatc 9300
atgaacaaag tcatcggaga gtttctgtcc aaccagcagc catacccccа gttcatggcc 9360
accgtggtgt ataaggtgtt tcagactctg cacagcaccg ggcagtcgtc catggtccgg 9420

gactgggtca tgctgtccct ctccaacttc acgcagaggg ccccggtcgc catggccacg 9480
tggagcctct cctgcttctt tgtcagcgcg tccaccagcc cgtgggtcgc ggcgatcctc 9540
ccacatgtca tcagcaggat gggcaagctg gagcaggtgg acgtgaacct tttctgcctg 9600
gtcgccacag acttctacag acaccagata gaggaggagc tcgaccgcag ggccttccag 9660
tctgtgcttg aggtggttgc agccccagga agcccatatc accggctgct gacttgttta 9720
cgaaatgtcc acaaggtcac cacctgctga gcgccatggt gggagagact gtgaggcggc 9780
agctggggcc ggagcctttg gaagtctgtg cccttgtgcc ctgcctccac cgagccagct 9840
tggtccctat gggcttccgc acatgccgcg ggcggccagg caacgtgcgt gtctctgcca 9900
tgtggcagaa gtgctctttg tggcagtggc caggcaggga gtgtctgcag tcctggtggg 9960
gctgagcctg aggccttcca gaaagcagga gcagctgtgc tgcaccccat gtgggtgacc 10020
aggtcctttc tcctgatagt cacctgctgg ttgttgccag gttgcagctg ctcttgcatc 10080
tgggccagaa gtcctccctc ctgcaggctg gctgttggcc cctctgctgt cctgcagtag 10140
aaggtgccgt gagcaggctt tgggaacact ggcctgggtc tccctggtgg ggtgtgcatg 10200
ccacgccccg tgtctggatg cacagatgcc atggcctgtg ctgggccagt ggctgggggt 10260
gctagacacc cggcaccatt ctcccttctc tcttttcttc tcaggattta aaatttaatt 10320
atatcagtaa agagattaat tttaacgaac tctttctatg cccgtgtaaa gtatgtgaat 10380
cgcaaggcct gtgctgcatg cgacagcgtc cggggtggtg gacagggccc ccggccacgc 10440
tccctctcct gtagccactg gcatagccct cctgagcacc cgctgacatt tccgttgtac 10500
atgttcctgt ttatgcattc acaaggtgac tgggatgtag agaggcgtta gtgggcaggt 10560
ggccacagca ggactgagga caggcccccа ttatcctagg ggtgcgctca actgcagccc 10620
ctcctcctcg ggcacagacg actgtcgttc tccacccacc agtcagggac agcagcctcc 10680
ctgtcactca gctgagaagg ccagccctcc ctggctgtga gcagcctcca ctgtgtccag 10740
agacatgggc ctcccactcc tgttccttgc tagccctggg gtggcgtctg cctaggagct 10800
ggctggcagg tgttgggacc tgctgctcca tggatgcatg ccctaagagt gtcactgagc 10860
tgtgttttgt ctgagcctct ctcggtcaac agcaaagctt ggtgtcttgg cactgttagt 10920
gacagagccc agcatccctt ctgcccccgt tccagctgac atcttgcacg gtgacccctt 10980
ttagtcagga gagtgcagat ctgtgctcat cggagactgc cccacggccc tgtcagagcc 11040
gccactccta tccccaggac aggtccctgg accagcctcc tgtttgcagg cccagaggag 11100
ccaagtcatt aaaatggaag tggattctgg atggccgggc tgctgctgat gtaggagctg 11160
gatttgggag ctctgcttgc cgactggctg tgagacgagg caggggctct gcttcctcag 11220
ccctagaggc gagccaggca aggttggcga ctgtcatgtg gcttggtttg gtcatgcccg 11280
tcgatgtttt gggtattgaa tgtggtaagt ggaggaaatg ttggaactct gtgcaggtgc 11340
tgccttgaga cccccaagct tccacctgtc cctctcctat gtggcagctg gggagcagct 11400
gagatgtgga cttgtatgct gcccacatac gtgaggggga gctgaaaggg agcccctgct 11460
caaagggagc ccctcctctg agcagcctct gccaggcctg tatgaggctt ttcccaccag 11520
ctcccaacag aggcctcccc cagccaggac cacctcgtcc tcgtggcggg gcagcaggag 11580
cggtagaaag gggtccgatg tttgaggagg cccttaaggg aagctactga attataacac 11640
gtaagaaaat caccattctt ccgtattggt tgggggctcc tgtttctcat cctagctttt 11700
tcctggaaaa gcccgctaga aggtttggga acgaggggaa agttctcaga actgttgctg 11760
ctccccaccc gcctcccgcc tcccccgcag gttatgtcag cagctctgag acagcagtat 11820
cacaggccag atgttgttcc tggctagatg tttacatttg taagaaataa cactgtgaat 11880
gtaaaacaga gccattccct tggaatgcat atcgctgggc tcaacataga gtttgtcttc 11940
ctcttgttta cgacgtgatc taaaccagtc cttagcaagg ggctcagaac accccgctct 12000
ggcagtaggt gtcccccacc cccaaagacc tgcctgtgtg ctccggagat gaatatgagc 12060
tcattagtaa aaatgacttc acccacgcat atacataaag tatccatgca tgtgcatata 12120
gacacatcta taattttaca cacacacctc tcaagacgga gatgcatggc ctctaagagt 12180
gcccgtgtcg gttcttcctg gaagttgact ttccttagac ccgccaggtc aagttagccg 12240
cgtgacggac atccaggcgt gggacgtggt cagggcaggg ctcattcatt gcccactagg 12300
atcccactgg cgaagatggt ctccatatca gctctctgca gaagggagga agactttatc 12360
atgttcctaa aaatctgtgg caagcaccca tcgtattatc caaattttgt tgcaaatgtg 12420
attaatttgg ttgtcaagtt ttgggggtgg gctgtgggga gattgctttt gttttcctgc 12480
tggtaatatc gggaaagatt ttaatgaaac cagggtagaa ttgtttggca atgcactgaa 12540
gcgtgtttct ttcccaaaat gtgcctccct tccgctgcgg gcccagctga gtctatgtag 12600
gtgatgtttc cagctgccaa gtgctctttg ttactgtcca ccctcatttc tgccagcgca 12660
tgtgtccttt caaggggaaa atgtgaagct gaacccccctc cagacaccca gaatgtagca 12720
tctgagaagg ccctgtgccc taaaggacac ccctcgcccc catcttcatg gaggggggtca 12780
tttcagagcc ctcggagcca atgaacagct cctcctcttg gagctgagat gagccccacg 12840
tggagctcgg gacggatagt agacagcaat aactcggtgt gtggccgcct ggcaggtgga 12900
acttcctccc gttgcggggt ggagtgaggt tagttctgtg tgtctggtgg gtggagtcag 12960
gcttctcttg ctacctgtga gcatccttcc cagcagacat cctcatcggg ctttgtccct 13020
ccccgcttc ctccctctgc ggggaggacc cgggaccaca gctgctggcc agggtagact 13080

```
tggagctgtc ctccagaggg gtcacgtgta ggagtgagaa gaaggaagat cttgagagct 13140
gctgagggac cttggagagc tcaggatggc tcagacgagg acactcgctt gccgggcctg 13200
gccctcctgg gaaggaggga gctgctcaga atgccgcatg acaactgaag gcaacctgga 13260
aggttcaggg cccgctcttc ccccatgtgc ctgtcacgct ctggtgcagt caaaggaacg 13320
ccttcccctc agttgtttct aagagcagag tctcccgctg caatctgggt ggtaactgcc 13380
agccttggag gatcgtggcc aacgtggacc tgcctacgga gggtgggctc tgacccaagt 13440
ggggcctcct tgcccaggtc tcactgcttt gcaccgtggt cagagggact gtcagctgag 13500
cttgagctcc cctggagcca gcagggctgt gatgggcgag tcccggagcc ccacccagac 13560
ctgaatgctt ctgagagcaa agggaaggac tgacgagaga tgtatattta atttttaac 13620
tgctgcaaac attgtacatc caaattaaag ggaaaaaatg gaaaccatca at         13672
```

<210> 2
<211> 3144
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Ala Thr Leu Glu Lys Leu Met Lys Ala Phe Glu Ser Leu Lys Ser
 1               5                  10                  15
Phe Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
            20                  25                  30
Gln Gln Gln Gln Gln Gln Gln Pro Pro Pro Pro Pro Pro Pro Pro Pro
        35                  40                  45
Pro Pro Pro Gln Leu Pro Gln Pro Pro Pro Gln Ala Gln Pro Leu Leu
    50                  55                  60
Pro Gln Pro Gln Pro Pro Pro Pro Pro Pro Pro Pro Pro Gly Pro
65                  70                  75                  80
Ala Val Ala Glu Glu Pro Leu His Arg Pro Lys Lys Glu Leu Ser Ala
                85                  90                  95
Thr Lys Lys Asp Arg Val Asn His Cys Leu Thr Ile Cys Glu Asn Ile
            100                 105                 110
Val Ala Gln Ser Val Arg Asn Ser Pro Glu Phe Gln Lys Leu Leu Gly
        115                 120                 125
Ile Ala Met Glu Leu Phe Leu Leu Cys Ser Asp Asp Ala Glu Ser Asp
    130                 135                 140
Val Arg Met Val Ala Asp Glu Cys Leu Asn Lys Val Ile Lys Ala Leu
145                 150                 155                 160
Met Asp Ser Asn Leu Pro Arg Leu Gln Leu Glu Leu Tyr Lys Glu Ile
                165                 170                 175
Lys Lys Asn Gly Ala Pro Arg Ser Leu Arg Ala Ala Leu Trp Arg Phe
            180                 185                 190
Ala Glu Leu Ala His Leu Val Arg Pro Gln Lys Cys Arg Pro Tyr Leu
        195                 200                 205
Val Asn Leu Leu Pro Cys Leu Thr Arg Thr Ser Lys Arg Pro Glu Glu
    210                 215                 220
Ser Val Gln Glu Thr Leu Ala Ala Ala Val Pro Lys Ile Met Ala Ser
225                 230                 235                 240
Phe Gly Asn Phe Ala Asn Asp Asn Glu Ile Lys Val Leu Leu Lys Ala
                245                 250                 255
Phe Ile Ala Asn Leu Lys Ser Ser Ser Pro Thr Ile Arg Arg Thr Ala
            260                 265                 270
Ala Gly Ser Ala Val Ser Ile Cys Gln His Ser Arg Arg Thr Gln Tyr
        275                 280                 285
Phe Tyr Ser Trp Leu Leu Asn Val Leu Leu Gly Leu Leu Val Pro Val
    290                 295                 300
Glu Asp Glu His Ser Thr Leu Leu Ile Leu Gly Val Leu Leu Thr Leu
305                 310                 315                 320
Arg Tyr Leu Val Pro Leu Leu Gln Gln Gln Val Lys Asp Thr Ser Leu
                325                 330                 335
```

```
Lys Gly Ser Phe Gly Val Thr Arg Lys Glu Met Glu Val Ser Pro Ser
            340                 345                 350
Ala Glu Gln Leu Val Gln Val Tyr Glu Leu Thr Leu His His Thr Gln
        355                 360                 365
His Gln Asp His Asn Val Val Thr Gly Ala Leu Glu Leu Leu Gln Gln
    370                 375                 380
Leu Phe Arg Thr Pro Pro Pro Glu Leu Leu Gln Thr Leu Thr Ala Val
385                 390                 395                 400
Gly Gly Ile Gly Gln Leu Thr Ala Ala Lys Glu Glu Ser Gly Gly Arg
                405                 410                 415
Ser Arg Ser Gly Ser Ile Val Glu Leu Ile Ala Gly Gly Gly Ser Ser
            420                 425                 430
Cys Ser Pro Val Leu Ser Arg Lys Gln Lys Gly Lys Val Leu Leu Gly
        435                 440                 445
Glu Glu Glu Ala Leu Glu Asp Asp Ser Glu Ser Arg Ser Asp Val Ser
    450                 455                 460
Ser Ser Ala Leu Thr Ala Ser Val Lys Asp Glu Ile Ser Gly Glu Leu
465                 470                 475                 480
Ala Ala Ser Ser Gly Val Ser Thr Pro Gly Ser Ala Gly His Asp Ile
                485                 490                 495
Ile Thr Glu Gln Pro Arg Ser Gln His Thr Leu Gln Ala Asp Ser Val
            500                 505                 510
Asp Leu Ala Ser Cys Asp Leu Thr Ser Ser Ala Thr Asp Gly Asp Glu
        515                 520                 525
Glu Asp Ile Leu Ser His Ser Ser Ser Gln Val Ser Ala Val Pro Ser
    530                 535                 540
Asp Pro Ala Met Asp Leu Asn Asp Gly Thr Gln Ala Ser Ser Pro Ile
545                 550                 555                 560
Ser Asp Ser Ser Gln Thr Thr Thr Glu Gly Pro Asp Ser Ala Val Thr
                565                 570                 575
Pro Ser Asp Ser Ser Glu Ile Val Leu Asp Gly Thr Asp Asn Gln Tyr
            580                 585                 590
Leu Gly Leu Gln Ile Gly Gln Pro Gln Asp Glu Asp Glu Glu Ala Thr
        595                 600                 605
Gly Ile Leu Pro Asp Glu Ala Ser Glu Ala Phe Arg Asn Ser Ser Met
    610                 615                 620
Ala Leu Gln Gln Ala His Leu Leu Lys Asn Met Ser His Cys Arg Gln
625                 630                 635                 640
Pro Ser Asp Ser Ser Val Asp Lys Phe Val Leu Arg Asp Glu Ala Thr
                645                 650                 655
Glu Pro Gly Asp Gln Glu Asn Lys Pro Cys Arg Ile Lys Gly Asp Ile
            660                 665                 670
Gly Gln Ser Thr Asp Asp Asp Ser Ala Pro Leu Val His Cys Val Arg
        675                 680                 685
Leu Leu Ser Ala Ser Phe Leu Leu Thr Gly Gly Lys Asn Val Leu Val
    690                 695                 700
Pro Asp Arg Asp Val Arg Val Ser Val Lys Ala Leu Ala Leu Ser Cys
705                 710                 715                 720
Val Gly Ala Ala Val Ala Leu His Pro Glu Ser Phe Phe Ser Lys Leu
                725                 730                 735
Tyr Lys Val Pro Leu Asp Thr Thr Glu Tyr Pro Glu Glu Gln Tyr Val
            740                 745                 750
Ser Asp Ile Leu Asn Tyr Ile Asp His Gly Asp Pro Gln Val Arg Gly
        755                 760                 765
Ala Thr Ala Ile Leu Cys Gly Thr Leu Ile Cys Ser Ile Leu Ser Arg
    770                 775                 780
Ser Arg Phe His Val Gly Asp Trp Met Gly Thr Ile Arg Thr Leu Thr
785                 790                 795                 800
Gly Asn Thr Phe Ser Leu Ala Asp Cys Ile Pro Leu Leu Arg Lys Thr
                805                 810                 815
Leu Lys Asp Glu Ser Ser Val Thr Cys Lys Leu Ala Cys Thr Ala Val
```

```
            820                 825                 830
Arg Asn Cys Val Met Ser Leu Cys Ser Ser Ser Tyr Ser Glu Leu Gly
        835                 840                 845
Leu Gln Leu Ile Ile Asp Val Leu Thr Leu Arg Asn Ser Ser Tyr Trp
    850                 855                 860
Leu Val Arg Thr Glu Leu Leu Glu Thr Leu Ala Glu Ile Asp Phe Arg
865                 870                 875                 880
Leu Val Ser Phe Leu Glu Ala Lys Ala Glu Asn Leu His Arg Gly Ala
                885                 890                 895
His His Tyr Thr Gly Leu Leu Lys Leu Gln Glu Arg Val Leu Asn Asn
            900                 905                 910
Val Val Ile His Leu Leu Gly Asp Glu Asp Pro Arg Val Arg His Val
        915                 920                 925
Ala Ala Ala Ser Leu Ile Arg Leu Val Pro Lys Leu Phe Tyr Lys Cys
    930                 935                 940
Asp Gln Gly Gln Ala Asp Pro Val Val Ala Val Ala Arg Asp Gln Ser
945                 950                 955                 960
Ser Val Tyr Leu Lys Leu Leu Met His Glu Thr Gln Pro Pro Ser His
                965                 970                 975
Phe Ser Val Ser Thr Ile Thr Arg Ile Tyr Arg Gly Tyr Asn Leu Leu
            980                 985                 990
Pro Ser Ile Thr Asp Val Thr Met Glu Asn Asn Leu Ser Arg Val Ile
        995                 1000                1005
Ala Ala Val Ser His Glu Leu Ile Thr Ser Thr Thr Arg Ala Leu Thr
    1010                1015                1020
Phe Gly Cys Cys Glu Ala Leu Cys Leu Leu Ser Thr Ala Phe Pro Val
1025                1030                1035                1040
Cys Ile Trp Ser Leu Gly Trp His Cys Gly Val Pro Pro Leu Ser Ala
                1045                1050                1055
Ser Asp Glu Ser Arg Lys Ser Cys Thr Val Gly Met Ala Thr Met Ile
            1060                1065                1070
Leu Thr Leu Leu Ser Ser Ala Trp Phe Pro Leu Asp Leu Ser Ala His
        1075                1080                1085
Gln Asp Ala Leu Ile Leu Ala Gly Asn Leu Leu Ala Ala Ser Ala Pro
    1090                1095                1100
Lys Ser Leu Arg Ser Ser Trp Ala Ser Glu Glu Glu Ala Asn Pro Ala
1105                1110                1115                1120
Ala Thr Lys Gln Glu Glu Val Trp Pro Ala Leu Gly Asp Arg Ala Leu
                1125                1130                1135
Val Pro Met Val Glu Gln Leu Phe Ser His Leu Leu Lys Val Ile Asn
            1140                1145                1150
Ile Cys Ala His Val Leu Asp Asp Val Ala Pro Gly Pro Ala Ile Lys
        1155                1160                1165
Ala Ala Leu Pro Ser Leu Thr Asn Pro Pro Ser Leu Ser Pro Ile Arg
    1170                1175                1180
Arg Lys Gly Lys Glu Lys Glu Pro Gly Glu Gln Ala Ser Val Pro Leu
1185                1190                1195                1200
Ser Pro Lys Lys Gly Ser Glu Ala Ser Ala Ala Ser Arg Gln Ser Asp
                1205                1210                1215
Thr Ser Gly Pro Val Thr Thr Ser Lys Ser Ser Ser Leu Gly Ser Phe
            1220                1225                1230
Tyr His Leu Pro Ser Tyr Leu Lys Leu His Asp Val Leu Lys Ala Thr
        1235                1240                1245
His Ala Asn Tyr Lys Val Thr Leu Asp Leu Gln Asn Ser Thr Glu Lys
    1250                1255                1260
Phe Gly Gly Phe Leu Arg Ser Ala Leu Asp Val Leu Ser Gln Ile Leu
1265                1270                1275                1280
Glu Leu Ala Thr Leu Gln Asp Ile Gly Lys Cys Val Glu Glu Ile Leu
                1285                1290                1295
Gly Tyr Leu Lys Ser Cys Phe Ser Arg Glu Pro Met Met Ala Thr Val
            1300                1305                1310
```

Cys Val Gln Gln Leu Leu Lys Thr Leu Phe Gly Thr Asn Leu Ala Ser
1315 1320 1325
Gln Phe Asp Gly Leu Ser Ser Asn Pro Ser Lys Ser Gln Gly Arg Ala
1330 1335 1340
Gln Arg Leu Gly Ser Ser Ser Val Arg Pro Gly Leu Tyr His Tyr Cys
1345 1350 1355 1360
Phe Met Ala Pro Tyr Thr His Phe Thr Gln Ala Leu Ala Asp Ala Ser
1365 1370 1375
Leu Arg Asn Met Val Gln Ala Glu Gln Glu Asn Asp Thr Ser Gly Trp
1380 1385 1390
Phe Asp Val Leu Gln Lys Val Ser Thr Gln Leu Lys Thr Asn Leu Thr
1395 1400 1405
Ser Val Thr Lys Asn Arg Ala Asp Lys Asn Ala Ile His Asn His Ile
1410 1415 1420
Arg Leu Phe Glu Pro Leu Val Ile Lys Ala Leu Lys Gln Tyr Thr Thr
1425 1430 1435 1440
Thr Thr Cys Val Gln Leu Gln Lys Gln Val Leu Asp Leu Leu Ala Gln
1445 1450 1455
Leu Val Gln Leu Arg Val Asn Tyr Cys Leu Leu Asp Ser Asp Gln Val
1460 1465 1470
Phe Ile Gly Phe Val Leu Lys Gln Phe Glu Tyr Ile Glu Val Gly Gln
1475 1480 1485
Phe Arg Glu Ser Glu Ala Ile Ile Pro Asn Ile Phe Phe Phe Leu Val
1490 1495 1500
Leu Leu Ser Tyr Glu Arg Tyr His Ser Lys Gln Ile Ile Gly Ile Pro
1505 1510 1515 1520
Lys Ile Ile Gln Leu Cys Asp Gly Ile Met Ala Ser Gly Arg Lys Ala
1525 1530 1535
Val Thr His Ala Ile Pro Ala Leu Gln Pro Ile Val His Asp Leu Phe
1540 1545 1550
Val Leu Arg Gly Thr Asn Lys Ala Asp Ala Gly Lys Glu Leu Glu Thr
1555 1560 1565
Gln Lys Glu Val Val Val Ser Met Leu Leu Arg Leu Ile Gln Tyr His
1570 1575 1580
Gln Val Leu Glu Met Phe Ile Leu Val Leu Gln Gln Cys His Lys Glu
1585 1590 1595 1600
Asn Glu Asp Lys Trp Lys Arg Leu Ser Arg Gln Ile Ala Asp Ile Ile
1605 1610 1615
Leu Pro Met Leu Ala Lys Gln Gln Met His Ile Asp Ser His Glu Ala
1620 1625 1630
Leu Gly Val Leu Asn Thr Leu Phe Glu Ile Leu Ala Pro Ser Ser Leu
1635 1640 1645
Arg Pro Val Asp Met Leu Leu Arg Ser Met Phe Val Thr Pro Asn Thr
1650 1655 1660
Met Ala Ser Val Ser Thr Val Gln Leu Trp Ile Ser Gly Ile Leu Ala
1665 1670 1675 1680
Ile Leu Arg Val Leu Ile Ser Gln Ser Thr Glu Asp Ile Val Leu Ser
1685 1690 1695
Arg Ile Gln Glu Leu Ser Phe Ser Pro Tyr Leu Ile Ser Cys Thr Val
1700 1705 1710
Ile Asn Arg Leu Arg Asp Gly Asp Ser Thr Ser Thr Leu Glu Glu His
1715 1720 1725
Ser Glu Gly Lys Gln Ile Lys Asn Leu Pro Glu Glu Thr Phe Ser Arg
1730 1735 1740
Phe Leu Leu Gln Leu Val Gly Ile Leu Leu Glu Asp Ile Val Thr Lys
1745 1750 1755 1760
Gln Leu Lys Val Glu Met Ser Glu Gln Gln His Thr Phe Tyr Cys Gln
1765 1770 1775
Glu Leu Gly Thr Leu Leu Met Cys Leu Ile His Ile Phe Lys Ser Gly
1780 1785 1790
Met Phe Arg Arg Ile Thr Ala Ala Ala Thr Arg Leu Phe Arg Ser Asp

```
        1795                1800                1805
Gly Cys Gly Gly Ser Phe Tyr Thr Leu Asp Ser Leu Asn Leu Arg Ala
    1810                1815                1820
Arg Ser Met Ile Thr Thr His Pro Ala Leu Val Leu Leu Trp Cys Gln
1825                1830                1835                1840
Ile Leu Leu Leu Val Asn His Thr Asp Tyr Arg Trp Trp Ala Glu Val
                1845                1850                1855
Gln Gln Thr Pro Lys Arg His Ser Leu Ser Ser Thr Lys Leu Leu Ser
            1860                1865                1870
Pro Gln Met Ser Gly Glu Glu Glu Asp Ser Asp Leu Ala Ala Lys Leu
        1875                1880                1885
Gly Met Cys Asn Arg Glu Ile Val Arg Arg Gly Ala Leu Ile Leu Phe
    1890                1895                1900
Cys Asp Tyr Val Cys Gln Asn Leu His Asp Ser Glu His Leu Thr Trp
1905                1910                1915                1920
Leu Ile Val Asn His Ile Gln Asp Leu Ile Ser Leu Ser His Glu Pro
                1925                1930                1935
Pro Val Gln Asp Phe Ile Ser Ala Val His Arg Asn Ser Ala Ala Ser
            1940                1945                1950
Gly Leu Phe Ile Gln Ala Ile Gln Ser Arg Cys Glu Asn Leu Ser Thr
        1955                1960                1965
Pro Thr Met Leu Lys Lys Thr Leu Gln Cys Leu Glu Gly Ile His Leu
    1970                1975                1980
Ser Gln Ser Gly Ala Val Leu Thr Leu Tyr Val Asp Arg Leu Leu Cys
1985                1990                1995                2000
Thr Pro Phe Arg Val Leu Ala Arg Met Val Asp Ile Leu Ala Cys Arg
                2005                2010                2015
Arg Val Glu Met Leu Leu Ala Ala Asn Leu Gln Ser Ser Met Ala Gln
            2020                2025                2030
Leu Pro Met Glu Glu Leu Asn Arg Ile Gln Glu Tyr Leu Gln Ser Ser
        2035                2040                2045
Gly Leu Ala Gln Arg His Gln Arg Leu Tyr Ser Leu Leu Asp Arg Phe
    2050                2055                2060
Arg Leu Ser Thr Met Gln Asp Ser Leu Ser Pro Ser Pro Pro Val Ser
2065                2070                2075                2080
Ser His Pro Leu Asp Gly Asp Gly His Val Ser Leu Glu Thr Val Ser
                2085                2090                2095
Pro Asp Lys Asp Trp Tyr Val His Leu Val Lys Ser Gln Cys Trp Thr
            2100                2105                2110
Arg Ser Asp Ser Ala Leu Leu Glu Gly Ala Glu Leu Val Asn Arg Ile
        2115                2120                2125
Pro Ala Glu Asp Met Asn Ala Phe Met Met Asn Ser Glu Phe Asn Leu
    2130                2135                2140
Ser Leu Leu Ala Pro Cys Leu Ser Leu Gly Met Ser Glu Ile Ser Gly
2145                2150                2155                2160
Gly Gln Lys Ser Ala Leu Phe Glu Ala Ala Arg Glu Val Thr Leu Ala
                2165                2170                2175
Arg Val Ser Gly Thr Val Gln Gln Leu Pro Ala Val His His Val Phe
            2180                2185                2190
Gln Pro Glu Leu Pro Ala Glu Pro Ala Ala Tyr Trp Ser Lys Leu Asn
        2195                2200                2205
Asp Leu Phe Gly Asp Ala Ala Leu Tyr Gln Ser Leu Pro Thr Leu Ala
    2210                2215                2220
Arg Ala Leu Ala Gln Tyr Leu Val Val Val Ser Lys Leu Pro Ser His
2225                2230                2235                2240
Leu His Leu Pro Pro Glu Lys Glu Lys Asp Ile Val Lys Phe Val Val
                2245                2250                2255
Ala Thr Leu Glu Ala Leu Ser Trp His Leu Ile His Glu Gln Ile Pro
            2260                2265                2270
Leu Ser Leu Asp Leu Gln Ala Gly Leu Asp Cys Cys Cys Leu Ala Leu
        2275                2280                2285
```

Gln Leu Pro Gly Leu Trp Ser Val Val Ser Ser Thr Glu Phe Val Thr
2290 2295 2300
His Ala Cys Ser Leu Ile Tyr Cys Val His Phe Ile Leu Glu Ala Val
2305 2310 2315 2320
Ala Val Gln Pro Gly Glu Gln Leu Leu Ser Pro Glu Arg Arg Thr Asn
2325 2330 2335
Thr Pro Lys Ala Ile Ser Glu Glu Glu Glu Glu Val Asp Pro Asn Thr
2340 2345 2350
Gln Asn Pro Lys Tyr Ile Thr Ala Ala Cys Glu Met Val Ala Glu Met
2355 2360 2365
Val Glu Ser Leu Gln Ser Val Leu Ala Leu Gly His Lys Arg Asn Ser
2370 2375 2380
Gly Val Pro Ala Phe Leu Thr Pro Leu Leu Arg Asn Ile Ile Ile Ser
2385 2390 2395 2400
Leu Ala Arg Leu Pro Leu Val Asn Ser Tyr Thr Arg Val Pro Pro Leu
2405 2410 2415
Val Trp Lys Leu Gly Trp Ser Pro Lys Pro Gly Gly Asp Phe Gly Thr
2420 2425 2430
Ala Phe Pro Glu Ile Pro Val Glu Phe Leu Gln Glu Lys Glu Val Phe
2435 2440 2445
Lys Glu Phe Ile Tyr Arg Ile Asn Thr Leu Gly Trp Thr Ser Arg Thr
2450 2455 2460
Gln Phe Glu Glu Thr Trp Ala Thr Leu Leu Gly Val Leu Val Thr Gln
2465 2470 2475 2480
Pro Leu Val Met Glu Gln Glu Glu Ser Pro Pro Glu Glu Asp Thr Glu
2485 2490 2495
Arg Thr Gln Ile Asn Val Leu Ala Val Gln Ala Ile Thr Ser Leu Val
2500 2505 2510
Leu Ser Ala Met Thr Val Pro Val Ala Gly Asn Pro Ala Val Ser Cys
2515 2520 2525
Leu Glu Gln Gln Pro Arg Asn Lys Pro Leu Lys Ala Leu Asp Thr Arg
2530 2535 2540
Phe Gly Arg Lys Leu Ser Ile Ile Arg Gly Ile Val Glu Gln Glu Ile
2545 2550 2555 2560
Gln Ala Met Val Ser Lys Arg Glu Asn Ile Ala Thr His His Leu Tyr
2565 2570 2575
Gln Ala Trp Asp Pro Val Pro Ser Leu Ser Pro Ala Thr Thr Gly Ala
2580 2585 2590
Leu Ile Ser His Glu Lys Leu Leu Leu Gln Ile Asn Pro Glu Arg Glu
2595 2600 2605
Leu Gly Ser Met Ser Tyr Lys Leu Gly Gln Val Ser Ile His Ser Val
2610 2615 2620
Trp Leu Gly Asn Ser Ile Thr Pro Leu Arg Glu Glu Glu Trp Asp Glu
2625 2630 2635 2640
Glu Glu Glu Glu Glu Ala Asp Ala Pro Ala Pro Ser Ser Pro Pro Thr
2645 2650 2655
Ser Pro Val Asn Ser Arg Lys His Arg Ala Gly Val Asp Ile His Ser
2660 2665 2670
Cys Ser Gln Phe Leu Leu Glu Leu Tyr Ser Arg Trp Ile Leu Pro Ser
2675 2680 2685
Ser Ser Ala Arg Arg Thr Pro Ala Ile Leu Ile Ser Glu Val Val Arg
2690 2695 2700
Ser Leu Leu Val Val Ser Asp Leu Phe Thr Glu Arg Asn Gln Phe Glu
2705 2710 2715 2720
Leu Met Tyr Val Thr Leu Thr Glu Leu Arg Arg Val His Pro Ser Glu
2725 2730 2735
Asp Glu Ile Leu Ala Gln Tyr Leu Val Pro Ala Thr Cys Lys Ala Ala
2740 2745 2750
Ala Val Leu Gly Met Asp Lys Ala Val Ala Glu Pro Val Ser Arg Leu
2755 2760 2765
Leu Glu Ser Thr Leu Arg Ser Ser His Leu Pro Ser Arg Val Gly Ala

```
    2770                2775                2780
Leu His Gly Val Leu Tyr Val Leu Glu Cys Asp Leu Leu Asp Asp Thr
2785                2790                2795                2800
Ala Lys Gln Leu Ile Pro Val Ile Ser Asp Tyr Leu Leu Ser Asn Leu
                2805                2810                2815
Lys Gly Ile Ala His Cys Val Asn Ile His Ser Gln Gln His Val Leu
            2820                2825                2830
Val Met Cys Ala Thr Ala Phe Tyr Leu Ile Glu Asn Tyr Pro Leu Asp
        2835                2840                2845
Val Gly Pro Glu Phe Ser Ala Ser Ile Ile Gln Met Cys Gly Val Met
    2850                2855                2860
Leu Ser Gly Ser Glu Glu Ser Thr Pro Ser Ile Ile Tyr His Cys Ala
2865                2870                2875                2880
Leu Arg Gly Leu Glu Arg Leu Leu Leu Ser Glu Gln Leu Ser Arg Leu
                2885                2890                2895
Asp Ala Glu Ser Leu Val Lys Leu Ser Val Asp Arg Val Asn Val His
            2900                2905                2910
Ser Pro His Arg Ala Met Ala Ala Leu Gly Leu Met Leu Thr Cys Met
        2915                2920                2925
Tyr Thr Gly Lys Glu Lys Val Ser Pro Gly Arg Thr Ser Asp Pro Asn
    2930                2935                2940
Pro Ala Ala Pro Asp Ser Glu Ser Val Ile Val Ala Met Glu Arg Val
2945                2950                2955                2960
Ser Val Leu Phe Asp Arg Ile Arg Lys Gly Phe Pro Cys Glu Ala Arg
                2965                2970                2975
Val Val Ala Arg Ile Leu Pro Gln Phe Leu Asp Asp Phe Phe Pro Pro
            2980                2985                2990
Gln Asp Ile Met Asn Lys Val Ile Gly Glu Phe Leu Ser Asn Gln Gln
        2995                3000                3005
Pro Tyr Pro Gln Phe Met Ala Thr Val Val Tyr Lys Val Phe Gln Thr
    3010                3015                3020
Leu His Ser Thr Gly Gln Ser Ser Met Val Arg Asp Trp Val Met Leu
3025                3030                3035                3040
Ser Leu Ser Asn Phe Thr Gln Arg Ala Pro Val Ala Met Ala Thr Trp
                3045                3050                3055
Ser Leu Ser Cys Phe Phe Val Ser Ala Ser Thr Ser Pro Trp Val Ala
            3060                3065                3070
Ala Ile Leu Pro His Val Ile Ser Arg Met Gly Lys Leu Glu Gln Val
        3075                3080                3085
Asp Val Asn Leu Phe Cys Leu Val Ala Thr Asp Phe Tyr Arg His Gln
    3090                3095                3100
Ile Glu Glu Glu Leu Asp Arg Arg Ala Phe Gln Ser Val Leu Glu Val
3105                3110                3115                3120
Val Ala Ala Pro Gly Ser Pro Tyr His Arg Leu Leu Thr Cys Leu Arg
                3125                3130                3135
Asn Val His Lys Val Thr Thr Cys
            3140
```

<210> 3
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 3
ugcagcugau caucgaugug cugacccuga ggaacaguuc 40

<210> 4

<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 4
gaacuguucc ucagggucag cacaucgaug aucagcugca 40

<210> 5
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 5
tgtgctgact ctgaggaaca g 21

<210> 6
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 6
ugugcugacu cugaggaaca g 21

<210> 7
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 7
cuguuccuca gagucagcac a 21

<210> 8
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 8
catacctcaa actgcatgat g 21

<210> 9
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 9
cauaccucaa acugcaugau g 21

<210> 10
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 10
caucaugcag uuugagguau g 21

<210> 11
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 11
gcctgcagag ccggcggcct a 21

<210> 12
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 12
gccugcagag ccggcggccu a 21

<210> 13
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 13
uaggccgccg gcucugcagg c 21

<210> 14
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 14
acagagtttg tgacccacgc c 21

<210> 15
<211> 21

<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 15
acagaguuug ugacccacgc c 21

<210> 16
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 16
ggcguggguc acaaacucug u 21

<210> 17
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 17
tccctcatct actgtgtgca c 21

<210> 18
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 18
ucccucaucu acugugugca c 21

<210> 19
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 19
gugcacacag uagaugaggg a 21

<210> 20
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 20
ugugc 5

<210> 21
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 21
gcacauc 7

<210> 22
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 22
guugc 5

<210> 23
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 23
ggaagag 7

<210> 24
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 24
ugugcugacc cugaggaaca g 21

<210> 25
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 25
guuccucagg gucagcacau c 21

<210> 26
<211> 21
<212> DNA

<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 26
ugugcugacc cugaggaaaa g 21

<210> 27
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 27
uuuccucagg gucagcacau c 21

<210> 28
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 28
ugugcugacc cugaggaaaa g 21

<210> 29
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 29
guuccucagg gucagcacau c 21

<210> 30
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 30
gcgtaatacg actcactata ggaacagtat gtctcagaca tc 42

<210> 31
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 31

uucgaaguau uccgcguacg u 21

<210> 32
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 32
gcgtaatacg actcactata ggacaagcct aattagtgat gc 42

<210> 33
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic construct

<400> 33
gaacagtatg tctcagacat c 21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US04/29968

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : A61K 48/00; C07H 21/04
US CL : 424/93.1; 514/44; 536/24.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 424/93.1; 514/44; 536/24.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MILLER et al. Allele-specific silencing of dominant disease genes, PNAS, June 2003, Vol. 100, No. 12, pages 7195-7200.	1-28
X	FLUITER et al. Killing cancer by targeting genes that cancer cells have lost: allele-specific inhibition, a novel approach to the treatment of genetic disorders. Cell Mol Life Sci. May 2003, Vol. 60, No. 5, pages 834-843.	1-28
X	US 2003/0144239 A (Agami et al.) 30 July 2003 (30.07.2003), especially page 9 and claims	1-28

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
02 February 2005 (02.02.2005)

Date of mailing of the international search report
01 MAR 2005

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703) 305-3230

Authorized officer
Amy H Bowman
Telephone No. (571)272-0755

Form PCT/ISA/210 (second sheet) (January 2004)